Neandertal behaviour, diet, and disease inferred using ancient DNA preserved in dental calculus

Authors: Laura S Weyrich[1], Sebastian Duchene[2], Julien Soubrier[1], Luis Arriola[1], Bastien Llamas[1], James Breen[1], Alan G Morris[3], Kurt W Alt[4], David Caramelli[5], Veit Dresely[6], Milly Farrell[7], Andrew G Farrer[1], Michael Francken[8], Neville Gully[9], Wolfgang Haak[1], Karen Hardy[10], Katerina Harvati[8], Petra Held[11], Edward C. Holmes[2], John Kaidonis[9], Carles Lalueza-Fox[12], Marco de la Rasilla[13], Antonio Rosas[14], Patrick Semal[15], Arkadiusz Soltysiak[16], Grant Townsend[9], Donatella Usai[17], Joachim Wahl[18], Daniel H. Huson[19], Keith Dobney[20, 21, 22], and Alan Cooper[1*]

Affiliations:

[1]Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia

[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, University of Sydney, Sydney, Australia

[3]Department of Human Biology, University of Cape Town, Cape Town, South Africa

[4]Danube Private University, Krems, Austria; State Office for Heritage Management and Archaeology, Saxony-Anhalt, Germany; Heritage Museum, Halle, Germany; and Institute for Prehistory and Archaeological Science, Basel University, Switzerland

[5]Department of Biology, University of Florence, Florence, Italy

[6]Archaeology Saxony-Anhalt and Heritage Museum, Halle, Germany

[7]Human Origins and Palaeo Environments Group, Oxford Brookes University, Oxford, United Kingdom

[8]Paleoanthropology, Senckenberg Centre for Human Evolution and Paleoenvironments, Eberhard Karls University of Tübingen

[9]School of Dentistry, The University of Adelaide, Adelaide, Australia

[10]ICREA (Catalan Institution for Research and Advanced Studies) Pg. Lluís Companys 23, 08010 Barcelona, Catalonia, Spain; Departament de Prehistòria, Facultat de Filosofia i Lletres, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain.

[11]Institute of Anthropology, University of Mainz

[12]Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain

[13]Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain

[14]Paleoanthropology Group, Department of Paleobiology, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

[15]Scientific Service Heritage, Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[16]Department of Bioarchaeology, Institute of Archaeology, University of Warsaw, Warsaw, Poland.

[17]Istituto Italiano per l'Africa e l'Oriente (IsIAO), Rome, Italy

[18]State Office for Cultural Heritage Management Baden-Württemberg, Esslingen Germany

[19]Department of Algorithms in Bioinformatics, University of Tübingen, Tübingen, Germany

[20]Department of Archaeology, Classics and Egyptology, School of Histories, Languages and Cultures, University of Liverpool, Liverpool, United Kingdom

[21]Department of Archaeology, University of Aberdeen, Aberdeen, Scotland

[22]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia, Canada

Introductory Paragraph

Recent genomic data has revealed multiple interactions between Neandertals and modern humans[1], but there is currently little genetic evidence about Neandertal behavior, diet, or disease. We shotgun sequenced ancient DNA from five Neandertal dental calculus specimens to characterize regional differences in Neandertal ecology. At Spy cave, Belgium, Neandertal diet was heavily meat based, and included woolly rhinoceros and wild sheep (mouflon), characteristic of a steppe environment. In contrast to Spy, no meat was detected in the calculus specimens from El Sidrón cave, Spain - with dietary components of mushrooms, pine nuts, and moss reflecting forest gathering[2,3]. Differences in diet were also linked to an overall shift in the oral bacterial community (microbiota), and suggested that meat consumption contributed to significant variation between Neandertal microbiota. Evidence for self-medication was detected in an El Sidrón Neandertal with a dental abscess[4], who also suffered from a chronic gastrointestinal pathogen (*Enterocytozoon bieneusi*). Metagenomic data from this individual also contained a nearly complete genome of the archaeal commensal *Methanobrevibacter oralis* (10.2x depth of coverage) – the oldest draft microbial genome generated to date at ~48,000 years old. DNA preserved within dental calculus represents an important new resource of behavioral and health information for ancient hominin specimens, as well as a unique deep-time study system for microbial evolution.

Main Text

Neandertals remain our closest known extinct hominin relative that co-existed and occasionally interbred with anatomically modern humans (AMHs) across Eurasia in the Late Pleistocene[1]. Neandertals went extinct in Europe some 40,000 years ago (40 Kyr), while their demise across the rest of Eurasia is less clear[5]. Archaeological and isotopic data through the last glacial cycle (~120-12 Kyr) suggest that Neandertals were as carnivorous as polar bears or wolves[6] with a diet heavily based on large terrestrial herbivores, such as reindeer, woolly mammoth, and woolly rhinoceros[7]. In contrast, microwear analysis of tooth surfaces from Neandertals in different ecological settings, such as wooded areas or open plains, suggests that diets were guided by local food availability[3]. Phytolith, starch granule, and protein analysis from calcified dental plaque (calculus) indicate that Neandertal diet also included many plants, including some that were used for medicinal purposes[8]. As a result, what Neandertals ate remains a topic of considerable debate, with limited data on the specific animals and plants directly consumed or the potential impacts on Neandertal health and disease.

While genomic studies continue to reveal evidence of interbreeding between AMHs and Neandertals across Eurasia[9], little is known about the health consequences of these interactions. The genetic analysis of Neandertal dental calculus represents an opportunity to examine this issue and to reconstruct Neandertal diet, behavior, and disease[10]. Here, we report the first genetic analysis of dental calculus from five Neandertals (2 from Spy cave in Belgium; 2 from El Sidrón cave in Spain; and 1 from Breuil Grotta in Italy) and compare this to a historic chimpanzee (n=1) and a modern human (n=1), as well as low coverage sequencing of calculus from a wide-range of ancient humans (Table S1). To provide increased resolution of the diseases that may

100 have impacted Neandertals, we also deeply sequenced (>147 million reads) dental
101 calculus from the best preserved Neandertal (El Sidrón 1), which suffered from a
102 dental abscess[4].

103 Size-based PCR amplification biases can confound standard metabarcoding
104 analyses (*i.e.* 16S rRNA amplicon sequencing[11,12]) of ancient dental calculus[13].
105 Consequently, we compared metagenomic shotgun and 16S rRNA amplicon (V4
106 region) analyses of the Neandertal dental calculus specimens – the oldest examined to
107 date. The 16S amplicon data sets were not representative of the biodiversity revealed
108 by shotgun sequencing (Extended Data Figure 1 and 2; Tables S2, S3, and S7),
109 clustered together (Figure 1), and contained disproportionately large amounts of non-
110 oral and environmental contaminant microorganisms (Tables S2 and S7; Extended
111 Data Figure 3 and 4A). As a result, the 16S amplicon data sets were excluded from
112 downstream analysis, along with the Breuil Grotta Neandertal sample that failed to
113 produce amplifiable sequences.

114 The shotgun data sets consisted of short DNA fragments (*e.g.* <70 bp), which
115 complicated accurate bacterial species identification using standard software, such as
116 MG-RAST or DIAMOND (Extended Data Figure 4B)[14,15]. To circumvent this
117 problem, we benchmarked and used MALTX, a new metagenomic alignment tool that
118 rapidly identifies species from shorter fragment lengths using a rapid BLASTX-like
119 algorithm[15,16] (Extended Data Figure 5; Table S4, S5, S6 and S7). Bioinformatic
120 filtering revealed that the Spy Neandertals were more heavily impacted by
121 environmental contamination (Extended Data Figure 6; Table S7)[15]. Indeed, shotgun
122 sequences from Spy I presented DNA damage patterns characteristic of contamination
123 with modern DNA sequences (Extended Data Figure 7), clustered more closely to the
124 modern individual than other Neandertals (Figure 1), and contained similar diversity

125 to environmental samples (Extended Data Figure 8). Therefore, this individual was
126 also excluded from further analyses[15]. The three remaining robust Neandertal shotgun
127 data sets (El Sidrón 1, El Sidrón 2, and Spy II) contained on average 93.76%
128 bacterial, 5.91% archaeal, 0.27% eukaryotic, and 0.06% viral identifiable sequences,
129 similar to previously published ancient and modern human dental calculus (Figure 2A
130 and Extended Data Figure 6C)[12]. The six dominant bacterial phyla in the modern
131 human mouth (Actinobacteria, Firmicutes, Bacteroidetes, Fusobacteria,
132 Proteobacteria, and Spirochaetes) were also dominant in each of the Neandertals, with
133 an average of 222 bacterial species per individual (Figure 2A and Extended Data
134 Figure 6C)[15].
135 We first examined Neandertal diets using the eukaryotic diversity preserved
136 within the dental calculus[15]. Calculus from the Spy II individual contained high
137 numbers of reads mapping to rhinoceros (*Ceratotherium simum*) and sheep (*Ovis*
138 *aries*), as well as the edible 'grey shag' mushroom (*Coprinopsis cinerea*) (Table 1).
139 Bones of woolly rhino, reindeer, mammoth, and horses were present in Spy Cave[17],
140 while wild mouflon sheep were broadly distributed in Europe throughout the
141 Pleistocene[15,18]. Woolly rhino has long been suspected to be part of the Spy
142 Neandertal diet[19], confirming the highly carnivorous lifestyle inferred from the
143 isotope and dental microwear data obtained from the Spy individuals[3,6,20]. These
144 findings also support recent isotope evidence that suggests Spy Neandertals were
145 regularly consuming mushrooms[21].
146 The dietary profile of El Sidrón Neandertals was markedly different from Spy,
147 and contained no sequences matching large herbivores or suggesting high meat
148 consumption. However, reads mapping to edible mushrooms ('split gill';
149 *Schizophyllum commune*), pine nuts (*Pinus koraiensis*), forest moss (*Physcomitrella*

150 *patens*), and poplar (*Populus trichocarpa*) were identified (Table 1). Sequences
151 mapping to plant fungal pathogens were also observed (*Zymoseptoria tritici,*
152 *Phaeosphaeria nodorum, Penicillium rubens,* and *Myceliophthora thermophila*),
153 suggesting the El Sidrón Neandertals may have consumed molded herbaceous
154 material. Limited zooarchaeological evidence exists for the El Sidrón individuals, and
155 our first genetic description of their diet supports evidence that Neandertal groups
156 across Europe used multiple subsistence strategies according to location and food
157 availability[2,3]. Additional approaches are needed to verify and extend these dietary
158 reconstructions and explore the limitations of the current approach[18].
159 Our findings support previous suggestions that El Sidrón 1 may have been
160 self-medicating a dental abscess,[8] as this was the only individual whose calculus
161 contained sequences corresponding to poplar, which contains the natural pain killer
162 salicylic acid (*i.e.* the active ingredient in aspirin), and the natural antibiotic producing
163 *Penicillium*. This individual also yielded sequences matching the intracellular
164 eukaryotic pathogen microsporidia (*Enterocytozoon bieneusi*) that causes acute
165 diarrhea in humans[22], highlighting another health issue potentially needing self-
166 medication.
167 To examine if oral microorganisms (microbiota) in Neandertals reflected
168 dietary composition, we compared the filtered shotgun data to a wide range of ancient
169 calculus specimens from humans with varying diets, including ancient Later Stone
170 Age (LSA) African foragers; African Pastoralist Period individuals with high meat
171 consumption[23]; European hunter-gatherers with a diet that included a wide range of
172 protein sources; and early European farmers with diets largely based around high
173 carbohydrate and milk consumption[15]. We used UPGMA clustered Bray Curtis
174 distances[15] to reveal four distinct groups: forager-gatherers with limited meat

consumption (El Sidrón Neandertals, chimpanzee, and LSA African gatherers); hunter-gatherers (or pastoralists) with a frequent meat diet (Spy Neandertal, African pastoralists, and European hunter-gatherers); ancient agriculturalists (European farming individuals); and modern humans (Figure 2B; Extended Data Figure 9A). This analysis identifies a split between hunter-gatherers and agriculturalists, as previously observed[11], but also reveals two distinct hunter-gatherer groups, apparently differentiated by the quantity of meat consumed in their diet. Meat consumption appears to have impacted early hominin microbiota, analogous to carnivorous and herbivorous mammals[24]. This also suggests that microorganisms preserved in dental calculus can be used to record details of dietary behavior in ancient hominins.

We also examined the Neandertal microbial diversity for potential pathogens as a sign of disease. Neandertal microbiota were more similar to the historic chimpanzee sample than the modern human and contained less potentially pathogenic Gram-negative species, which are associated with secondary enamel colonization, increased plaque formation, and periodontal disease (18.9% Gram-negatives in Neandertals compared to 77.6% in the modern human; Extended Data Figure 9B)[25]. All types of microbial taxa were equally damaged and fragmented, suggesting this difference is not simply due to a preservation bias in Gram-negatives, as previously reported[11] (Table 2)[15]. The low levels of immunostimulatory Gram-negative taxa in Neandertals may be related to the reduced presence of Fusobacteria (Extended Data Figure 6C), which can facilitate the binding of Gram-negative microorganisms to the primary colonizers that bind tooth enamel (*e.g. Streptococcus*, *Actinomyces*, and *Methanobrevibacter* species)[26]. Notably, the increased diversity of Gram-negative immunostimulatory taxa in modern humans are strongly linked to a wide-range of Western diseases[27].

Several oral pathogens could be identified within the shotgun data, although the short ancient sequences and diverse metagenomic background made authentication complex. We established a number of criteria to verify the presence of specific bacterial pathogens, including the assessment of ancient DNA damage, phylogenetic position, and bioinformatic comparisons to close relatives[15]. We identified the caries-associated species *Streptococcus mutans* (0.08% to 0.18%) and the members of the 'red complex' associated with modern periodontal disease (*Porphyromonas gingivalis*: 0-0.52%; *Tannerella forsythia* 0.05-2.4%; and *Treponema denticola* 0-1.87%) in concordance with evidence of dental caries and periodontal disease in Neandertals[28] (Tables S9-S11). A variety of other pathogens (*Bordetella parapertusssis, Pasteurella multocida, Neisseria gonorrhoeae*, *Streptococcus pyogenes*, and *Corynebacterium diptheriae*) were identified but could not be unambiguously distinguished from closely related commensal oral taxa (Extended Data Figure 10A; Table S9 and S14), highlighting the need for rigorous criteria when identifying pathogenic strains from ancient metagenomic data.

Lastly, we examined Neandertal commensal microorganisms and within the deeply sequenced El Sidrón 1 specimen we were able to recover draft genomes (>1x depth of coverage) for the eight most prevalent microbial species (Table 2). Of particular interest was a dominant archaeal species (14.7%; Extended Data Figure 6C) in El Sidrón 1 that was present in lower proportions in other Neandertals (1.4% and 1.2% in El Sidrón 2 and Spy II, respectively). The large differences in G/C content between bacteria and archaea facilitated efficient read mapping (Table 2) to the modern human-associated *Methanobrevibacter oralis* JMR01 strain. At ~48 Kyr[29], *Methanobrevibacter oralis* subsp. *neandertalensis* is the oldest draft microbial

224 genome to date (44.7% of 2.1 Mbp covered at a 10.3x depth of coverage; Table 2 and
225 Figure 3).

226 Date estimates using a strict molecular clock place the divergence between the
227 Neandertal and modern human *M. oralis* strains between 112-143 Kyr (95% highest
228 posterior density interval; mean date of 126 Kyr) (Figure 3B)[15]. As this is long after
229 the genomic divergence of Neandertals and modern humans (450-750 Kyr)[30], it is
230 likely that commensal microbial species were transferred between the two hosts
231 during subsequent interactions in the Near East[31]. Further genome comparisons
232 revealed 136 coding sequences in the modern human *M. oralis* that were putatively
233 absent in *M. oralis* subsp. *neandertalensis* (Table S15), including genes encoding
234 antiseptic resistance (*qacE*), maltose metabolism regulation (*sfsA*), and bacterial
235 immunity (CRISPR Cas2 and Cas6; Table S15), likely reflecting dietary and hygiene
236 differences between modern humans and Neandertals. A comparison of 375
237 translatable protein coding sequences between the Neandertal and modern human *M.*
238 *oralis* indicated that 58% were under strong purifying selection ($d_N/d_S<0.1$) (Table
239 3)[15]. Only 4% appeared to be under putative positive selection ($d_N/d_S>1$), including
240 regions for conjugal gene transfer (*i.e.* uptake of foreign or plasmid DNA; *traB*) and
241 DNA mismatch repair (*mutT*).

242 Preserved dental calculus represents a critical new resource of behavioral,
243 dietary, and health information for ancient hominin specimens, as well as a unique
244 long-term system to study how hundreds of different microbial species have evolved
245 and spread amongst hominins.

246 **References**

247 1. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328,**
248 710–722 (2010).
249 2. Fiorenza, L. *et al.* Molar Macrowear Reveals Neanderthal Eco-Geographic
250 Dietary Variation. *PLOS ONE* **6,** e14769 (2011).
251 3. Zaatari, S. E., Grine, F. E., Ungar, P. S. & Hublin, J.-J. Neandertal versus Modern
252 Human Dietary Responses to Climatic Fluctuations. *PLOS ONE* **11,** e0153277
253 (2016).
254 4. Rosas, A. *et al.* Paleobiology and comparative morphology of a late Neandertal
255 sample from El Sidrón, Asturias, Spain. *PNAS* **103,** 19266–19271 (2006).
256 5. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal
257 disappearance. *Nature* **512,** 306–309 (2014).
258 6. Richards, M. P. & Trinkaus, E. Isotopic evidence for the diets of European
259 Neanderthals and early modern humans. *PNAS* **106,** 16034–16039 (2009).
260 7. Bocherens, H., Drucker, D. G., Billiou, D., Patou-Mathis, M. & Vandermeersch,
261 B. Isotopic evidence for diet and subsistence pattern of the Saint-Césaire I
262 Neanderthal: review and use of a multi-source mixing model. *Journal of Human*
263 *Evolution* **49,** 71–87 (2005).
264 8. Hardy, K. *et al.* Neanderthal medics? Evidence for food, cooking, and medicinal
265 plants entrapped in dental calculus. *Naturwissenschaften* **99,** 617–626 (2012).
266 9. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal
267 ancestor. *Nature* **524,** 216–219 (2015).
268 10. Weyrich, L. S., Dobney, K. & Cooper, A. Ancient DNA analysis of dental
269 calculus. *Journal of Human Evolution* (2015). doi:10.1016/j.jhevol.2014.06.018

11. Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet* **45,** 450–455 (2013).
12. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* **46,** 336–344 (2014).
13. Ziesemer, K. A. *et al.* Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* **5,** 16498 (2015).
14. Meyer, F. *et al.* The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9,** 386 (2008).
15. Materials and methods are available as supplementary materials on *Nature* Online.
16. Herbig, A. *et al.* MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* 050559 (2016). doi:10.1101/050559
17. Germonpré, M., Udrescu, M. & Fiers, E. The fossil mammals of Spy. *Anthropologica et Præhistorica* **123/2012,** 298–327 (2013).
18. Nouvelles données paléogéographiques et chronologiques sur les Caprinae (Mammalia, Bovidae) du Pléistocène moyen et supérieur d'Europe :: EI-SEV publishing: :: Euskomedia. Available at: http://www.euskomedia.org/analitica/10263. (Accessed: 27th July 2016)
19. Patou-Mathis, M. Neanderthal subsistence behaviours in Europe. *International Journal of Osteoarchaeology* **10,** 379–395 (2000).

20. Naito, Y. I. *et al.* Ecological niche of Neanderthals from Spy Cave revealed by nitrogen isotopes of individual amino acids in collagen. *Journal of Human Evolution* **93,** 82–90 (2016).
21. O'Regan, H. J., Lamb, A. L. & Wilkinson, D. M. The missing mushrooms: Searching for fungi in ancient human dietary analysis. *Journal of Archaeological Science* doi:10.1016/j.jas.2016.09.009
22. Tumwine, J. K. *et al.* Enterocytozoon bieneusi among children with diarrhea attending Mulago Hospital in Uganda. *Am. J. Trop. Med. Hyg.* **67,** 299–303 (2002).
23. Leonard, W. R. & Crawford, M. H. *The Human Biology of Pastoral Populations*. (Cambridge University Press, 2002).
24. Muegge, B. D. *et al.* Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. *Science* **332,** 970–974 (2011).
25. Marsh, D. P. D. & Bradshaw, D. J. Dental plaque as a biofilm. *Journal of Industrial Microbiology* **15,** 169–175 (1995).
26. Signat, B., Roques, C., Poulet, P. & Duffaut, D. Fusobacterium nucleatum in periodontal health and disease. *Curr Issues Mol Biol* **13,** 25–36 (2011).
27. Cho, I. & Blaser, M. J. The Human Microbiome: at the interface of health and disease. *Nat Rev Genet* **13,** 260–270 (2012).
28. Topić, B., Raščić-Konjhodžić, H. & Cižek Sajko, M. Periodontal disease and dental caries from Krapina Neanderthal to contemporary man - skeletal studies. *Acta Med Acad* **41,** 119–130 (2012).
29. Wood, R. E. *et al.* A New Date for the Neanderthals from El Sidrón Cave (asturias, Northern Spain)*. *Archaeometry* **55,** 148–158 (2013).

30. Stringer, C. The origin and evolution of Homo sapiens. *Phil. Trans. R. Soc. B* **371,** 20150237 (2016).
31. Sankararaman, S., Patterson, N., Li, H., Paabo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS genetics* **8,** e1002947 (2012).

Figures and Legends

Figure 1: Comparison of 16S amplicon and shotgun data sets obtained from ancient, historic, and modern dental calculus samples. Filtered and unfiltered 16S rRNA amplicon and shotgun data sets, as well as 16S rRNA shotgun sequences identified using GraftM, were compared using UPGMA clustering of Bray Curtis distances from a chimpanzee (red), Neandertals (El Sidrón 1 (dark green), El Sidrón 2 (light green), Spy I (grey), Spy II (blue), and a modern human (orange) (n=6).

Figure 2: Bacterial community composition at the phyla level of oral microbiota from chimpanzee, Neandertal, and modern human samples. (A) - Oral microbiota from shotgun data sets of a wild-caught chimpanzee (n=1), Neandertals (n=3), and a modern human (n=1) are presented at the phyla level. Phyla names were simplified for clarity, and unidentified reads were excluded. Gram-positive (blue) and Gram-negative (red) phyla are differentiated by color. (B) UPGMA clustering of Bray Curtis values obtained from 22 oral metagenomes is displayed. Definitions for abbreviations can be found in the SI.

Figure 3: 48,000 year old archaeal draft genome and phylogeny of *Methanobrevibacter oralis neandertalensis*.

(A) Ancient sequences mapping to *Methanobrevibacter oralis* JMR01 are displayed in a Circos plot (black), alongside the depth of coverage obtained (red). The reference sequence is displayed (grey) with the GC content of the reference sequence calculated in 2500 bp bins (green). (B) A Methanobrevibacter phylogeny was constructed from whole genome alignments in RAxML with 100 bootstrap replicates, with the percent support shown in each node. The estimated dates placed onto this tree were calculated from a whole genome phylogeny using a Bayesian methodology (in BEAST) assuming a strict clock model.

Table 1: Dietary information preserved in calculus.

DNA sequences mapping to eukaryotic species are shown as a proportion of the total eukaryotic reads identified within each sample. Eukaryotic sequencing identified in the extraction blank controls and the Spy I Neandertal, which is heavily contaminated with modern DNA, are shown to the right. * denotes samples or taxa that are likely the results of contamination, as they do not represent biological processes[15].

Scientific Name	Common Name of Likely Source	Hominid Pathogen (X) or Medicinal Uses (+)	El Sidron 1	El Sidron 2	Spy II	Chimpanzee	Moder Humar
Zymoseptoria tritici	Plant (wheat) pathogen		4.13%	0	0	0	0
Phaeosphaeria nodorum	Plant (wheat) pathogen		12.22%	0	0	3.98%	0
Penicillium rubens	Food fungus	+	3.97%	0	0	0	0
Myceliophthora thermophila	Cellulose Fungus		0	0	0.56%	0	0
Coprinopsis cinerea	Edible Mushrom (grey shag)		0	0	2.44%	0	0
Schizophyllum commune	Edible Mushroom (split gill)		3.65%	0	0	0	0
Malassezia globosa	Human fungal commensal		3.65%	8.89%	0	0	19.92%
Enterocytozoon bieneusi	Intracellular parasite (microsporidia)	X	8.10%	0	0	0	0
Ovis orientalis	Sheep (wild mouflon)		0	0	62.03%	0	0
Ceratotherium simum	White Rhino (woolly rhino)		0	0	34.40%	0	0
*Ixodes scapularis**	Tick		0	0	0	0	2.15%
Physcomitrella patens	Moss		2.06%	0	0	0	0
Pinus koraiensis	Pine Tree		13.49%	19.60%	0	4.45%	0

Populus trichocarpa	Poplar Tree	+	2.86%	0	0	0	0
Total Eukaryotic Reads			630	551	532	427	3760

346

Table 2: Draft microbial genomes present in El Sidrón 1.

Eight draft microbial genomes from Gram-positive, Gram-negative, eubacterial, and archaea were obtained from the deeply sequenced El Sidrón 1 specimen by read mapping. The sequence coverage, GC content, sequencing depth and damage profile (average fragment length and base pair modifications calculated from MapDamage2) are displayed for each genome.

Reference Genome				Mapped Reads					
Reference Genome	Sequence Reference Number	Length (Mbps)	GC content (%)	Base Covered (Mbps)	Unique Hits	Depth (Avg coverage)	Average Read Length	5p-C-T	3p-G-A
Methanobrevibacter oralis JMR01	NZ_CBWS00000000	2.107	27.8	0.941	370115	15.16	58.67	0.33	0.36
Candidatus Saccharibacteria oral TM7	NZ_CP007496.1	0.705	44.5	0.131	108919	5.83	52.46	0.37	0.41
Campylobacter gracilis ATCC 33236	NZ_CP012196.1	2.282	46.6	1.199	94472	2.40	51.7	0.38	0.41
Propionibacterium propionicum F0230a	NZ_018142.1	3.449	66.1	2.083	130748	1.89	48.85	0.37	0.43
Fretibacterium fastidiosum	gi 296110870	2.728	55.5	1.466	121822	2.43	48	0.39	0.43
Eubacterium infirmum F0142	NZ_AGWI00000000	1.9	40.1	0.176	52170	10.73	51.53	0.33	0.38
Peptostreptococcus stomatis DSM 17678	GCF_000147675.1	1.988	36.7	1.222	94743	2.90	54.62	0.36	0.4
Eubacterium sphenum ATCC 49989	NZ_GG688422.1	1.084	40.6	0.261	23124	3.46	52.87	0.37	0.41

Table 3: Purifying and positive selection in *M. oralis neandertalensis.*

The ratio of non-synonymous to synonymous substitutions per site (d_N/d_S) was calculated for coding regions with sufficient coverage and that were conserved between *M. oralis* and *M. oralis* subsp. *neandertalensis.* Genes that have undergone strong purifying ($d_N/d_S < 0.1$) or positive ($d_N/d_S > 1$; grey) selection are displayed if the function of the gene was annotated. Hypothetical proteins and those not matching to the *M. oralis* genome during BLAST searches are not shown.

Gene Number	CDS	Genebank	dN/dS ratio	Gene Annotation	Coding Protein Function
Gene1211	1184	WP_042693702.1	0	NZ_HG796201.1	preprotein translocase subunit SecG
Gene291	283	WP_042691749.1	0	NZ_HG796199.1	SAM-dependent methyltransferase
Gene303	295	WP_042691777.1	0	NZ_HG796199.1	fibrillarin
Gene343	343	WP_042691868.1	0	NZ_HG796199.1	sugar fermentation stimulation protein SfsA
Gene394	394	WP_042691937.1	0	NZ_HG796199.1	30S ribosomal protein S2
Gene401	401	WP_042691950.1	0	NZ_HG796199.1	transcriptional regulator
Gene745	745	WP_042692741.1	0	NZ_HG796200.1	50S ribosomal protein L37
Gene757	757	WP_042693268.1	0	NZ_HG796200.1	acyltransferase
Gene766	766	WP_042692795.1	0	NZ_HG796200.1	DNA-directed RNA polymerase
Gene769	769	WP_042692805.1	0	NZ_HG796200.1	30S ribosomal protein S6
Gene772	772	WP_042692815.1	0	NZ_HG796200.1	50S ribosomal protein L24
Gene773	773	WP_042692817.1	0	NZ_HG796200.1	30S ribosomal protein
Gene810	810	WP_042692911.1	0	NZ_HG796200.1	transcriptional regulator
Gene836	836	WP_042692956.1	0	NZ_HG796200.1	endonuclease DDE
Gene880	880	WP_042693050.1	1.52	NZ_HG796200.1	uracil transporter

Gene724	724	WP_042692699.1	2.67	NZ_HG796200.1	acetylesterase
Gene269	269	WP_042691703.1	3.64	NZ_HG796199.1	conjugal transfer protein TraB
Gene1206	1206	WP_042693692.1	12	NZ_HG796201.1	DNA mismatch repair protein MutT

357

Acknowledgements

We thank Giorgio Manzi (University of Rome), the Odontological Collection of the Royal College of Surgeons, Royal Belgian Institute of Natural Sciences, Museo Nacional de Ciencias Naturales, and Adelaide Universities for access to dental calculus material. We thank Adam Croxford for DNA sequencing and Alan Walker, Johannes Krause, and Alexander Herbig for critical feedback. We dedicate this paper to the memory of Professor Don Brothwell (1933-2016) – who 'lit the fuse' of dental calculus studies in archaeology.

Author Contributions

LSW, KD, and AC designed study; AGM, KWA, DC, VD, MF, MF, NG, WH, KH, KH, PH, JK, CLF, MR, AR, PS, AS, DU, and JW provided samples and interpretations of associated archaeological goods; LSW performed experiments; LSW, SD, EH, JS, BL, JB, LA, and AGF performed bioinformatics analysis and interpretation of the data; DHH developed bioinformatics tools; NG, JK, and GT analyzed medical relevance of data; LSW, KD and AC wrote the paper; and all authors contributed to editing the manuscript.

Author Information

Raw and analyzed data sets and the scripts utilized for this analysis are available in the Online Ancient Genome Repository (OAGR). The Australian Research Council supported this work, and the authors declare no competing financial interests. Requests for additional materials can be addressed to laura.weyrich@adelaide.edu.au.

373 **Methods**

374 *Sampling handling and DNA extraction*

375 Samples were stored and all molecular biology procedures prior to PCR amplification stages were carried out at the Australian Centre for
376 Ancient DNA facility at The University of Adelaide. All experiments were performed within UV treated, still-air working hoods located in
377 isolated, still-air working rooms that have been designed to allow highly technical ancient DNA research to be performed with ultra-low levels
378 of background contamination (*i.e.* workflow is monitored, facilities are irradiated with ultraviolet light each night, and the general facility is
379 under positive air pressure). To minimize environmental contamination, each dental calculus sample was UV treated for 15 minutes on each
380 side, soaked in 2 mL of 5% bleach for 3 min, rinsed in 90% ethanol for 1 minute, and dried at room temperature for several minutes. Directly
381 proceeding decontamination, DNA extraction was performed using an in-house silica-based method, as previously described[33] but with
382 decreased buffer volumes (1.8 mL lysis buffer (1.6 mL EDTA; 200 uL SDS; 20 uL 20 mg/mL proteinase K) and 3 mL guanidine DNA binding
383 buffer).

384 *DNA library preparation and sequencing*

385 Once DNA was extracted, 16S ribosomal RNA amplicon libraries of the V4 region were constructed by PCR amplification[34]. Each
386 sample was amplified in triplicate, and samples were pooled, Ampure cleaned, and quantified using a TapeStation and quantitative PCR (KAPA
387 Illumina quantification kit), prior to sequencing with an Illumina MiSeq 300 cycle kit (~40 samples/run). Frequent and repetitive extraction

blank controls (EBCs) are used throughout all experimental procedures, *i.e.* extraction, amplification, and library preparation. Several key samples were selected for shotgun metagenomic sequencing. Shotgun metagenomic libraries were constructed as previously described[35], with 5 bp forward and reverse barcodes. Metagenomic libraries were Ampure cleaned, quantified using a TapeStation and quantitative PCR (KAPA Illumina quantification kit), and pooled at equimolar concentrations prior to sequencing.

16S rRNA amplicon library analysis

To process the 16S amplicon data, sequences were de-multiplexed using the CASAVA pipeline and joined into amplicons using fastq_joiner (ea-utils)[36]. Quality filtering and trimming was completed using Cutadapt, and sequences were then imported into QIIME 1.6.0 for analysis[37]. In QIIME, OTUs were clustered in UCLUST at 97%, and representative sequences were taxonomically identified using the Greengenes (gg_12_10) database[38]. After OTU selection, strict filtering was applied to all samples (SI Section II). Diversity was analysed in QIIME, and phylogenetic analysis was visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree). Statistical analysis were performed by anosim in QIIME, and the calculation of the Jaccard or Bray Curtis indices, hierarchal clustering, and heatmap construction was completed in R using the vegan and gplots packages (http://cran.r-project.org).

Shotgun DNA Library Analysis

To process shotgun metagenomic data, reads were merged with a 5 bp overlap using bbmerge, and reads matching the forward and reserve barcodes with one mismatch were retained using AdapterRemoval[39]. Taxonomic identifications were made using protein alignments in

MetaPhlAn [40], MG-RAST[41], DIAMOND[42], and the new Metagenome ALignment Tool with the BLASTX-like approach (MALTX) developed in the Huson laboratory at the University of Tubingen[16]. Taxonomic assignments were then filtered using default LCA parameters in MEGAN5[43], and data was exported at specific taxonomic classification levels (*i.e.* phyla, species, etc.) for downstream analysis. Reference genomes were excluded if they were known to have human DNA contamination[44]. Statistical analyses were done using a Mann-Whitney U test (comparisons of phyla in one taxa compared to other samples), a heteroscedastic t-test (direct comparisons between specific taxa in two samples/groups), or LefSe (identifying taxa that distinguish one group from another)[40]. Genomes were assembled by mapping to a reference genome using specific ancient DNA parameters in bwa[45], and authenticated using MapDamage 2.0[46]. Phylogenetic analyses were completed by mapping reads to reference genomes, aligning genomic sequences using progressiveMauve[47], and inferring trees in RAxML v8.1.21[48] using the GTRGAMMA model. A Bayesian approach was utilized to estimate dates of divergence between strains and clades. Detailed descriptions of each procedures are available in the Supplemental Information[49].

Data Availability Statement

Raw and analyzed sequence data and all analytical scripts used in this study are publically available in the Online Ancient Gene Repository (https://www.oagr.org.au/) under the study 'Reconstructing Neandertal behavior, diet, and disease using ancient DNA from dental calculus.' (doi: XXXXXXX).

418

References

1. Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* **4,** 1764 (2013).
2. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6,** 1621–1624 (2012).
3. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010,** pdb.prot5448 (2010).
4. Aronesty, E. *Command-line tools for processing biological sequencing data*. (2011).
5. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
6. DeSantis, T. Z. *et al.* Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* **72,** 5069–5072 (2006).
7. Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* **5,** 1–7 (2012).
8. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12,** R60 (2011).
9. Aziz RK *et al.* The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9,** (2008).
10. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Meth* **12,** 59–60 (2015).

433 11. Herbig, A. *et al.* MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* 050559
434 (2016). doi:10.1101/050559
435 12. Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4.
436 1552–1560 (2011).
437 13. Longo, M. S., O'Neill, M. J. & O'Neill, R. J. Abundant Human DNA Contamination Identified in Non-Primate Genome Databases. *PLOS*
438 *ONE* **6,** e16410 (2011).
439 14. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13,** 178 (2012).
440 15. Ginolhac, A., Rasmussen, M., Gilbert, M. T., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA
441 sequences. *Bioinformatics* **27,** 2153–5 (2011).
442 16. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS*
443 *ONE* **5,** e11147 (2010).
444 17. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.
445 *Bioinformatics* **22,** 2688–90 (2006).
446 18. Materials and methods are available as supplementary materials on *Nature* Online.
447
448

Extended Data Figure Legends

Extended Data Figure 1: Proportions of bacterial phyla from filtered and unfiltered 16S amplicon and shotgun datasets are plotted for El Sidron 1 (**A**) and the modern human oral microbiota (**B**). Samples in blue are from shotgun datasets, while red points are from 16S amplicon datasets. The different shapes of each data point correspond to the microbial phyla, which are displayed next to each phyla grouping (*e.g.* ✖ represents Proteobacteria for both 16S and shotgun data sets).

Extended Data Figure 2: As is consistent with ancient DNA metagenomic analysis, a presence/absence distance (Jaccard) were calculated for each OTU observed in the 99th percentile and clustered according to dissimilarity within each sample and amongst other OTUs. Clusters of unique OTUs are identified (dashed lines) and labelled according to cluster relationships (red: no-agriculture; green: agriculture; purple: 19th century; fuchsia: modern).

Extended Data Figure 3: SourceTracker take-one-out analysis was completed for all samples. Samples were grouped into time periods, the proportion of each taxa originating from each sample group was then inferred. "Other" represents summed proportions across non-oral microbial groups (non-oral human microbiome, air, and soil) and unknown classification. Groups have a minimum of two samples (Non-Human Primate group is removed in filtered analysis as filtering reduced sample number to one), and are displayed for the raw (unfiltered) OTU table data (A) (n = 54) and from filtered OTU table (B) (n = 42).

Extended Data Figure 4: (**A**) Unfiltered prokaryotic phyla identified from 16S rRNA results and shotgun sequencing results (MALTX) are compared. (**B**) Raw shotgun sequences were analyzed by MALTX and by MG-RAST, and bacterial phyla and kingdom level results are displayed.

Extended Data Figure 5: (**A**) Phyla identified in simulated metagenomes (modern or ancient) are shown for five different analysis programs: MALTX, DIAMOND, MetaPhylAn, and MG-RAST. (**B**) Simulated metagenomes (modern (circle) or ancient (square; damaged)) analyzed using four different software (DIAMOND (green); MALT (red); MetaPhylAn (blue); MG-RAST (orange)) were UPGMA clustered according to Bray Curtis distances calculated from genera within samples. (**C**) Phyla identified by MALTX analysis in shotgun and amplicon oral datasets obtained from this study and MG-RAST are displayed in stacked bar plots.

Extended Data Figure 6: (**A**) Phyla level identified sequenced from MALTX, are displayed for dental calculus samples, extraction blank controls (EBCs), and environmental samples. Ancient dental calculus samples are graphed in order of age, with the oldest specimens listed on the left. (**B**) Identified reads using MALTX were filtered to remove reads corresponding to species identified in extraction blank controls from QG DNA extractions and environmental controls. (**C**) Filtered data was summarized to examine only archaea and bacteria phyla typically found in the modern oral cavity. Dental calculus samples are displayed in order of age.

Extended Data Figure 7: MapDamage analysis was performed on the reads mapping to oral bacterial species that are shared between Neandertals and the modern human. The percent of C-T mutations (**A**) and read length (**B**) calculated from mapped reads of each sample is graphed for ten conserved species.

Extended Data Figure 8: Alpha diversity from deeply sequenced unfiltered shotgun datasets was calculated from rarefied data using Shannon-Weaver (A) and Simpson's Reciprocal (B) indexes.

Extended Data Figure 9: (**A**) UPGMA clustering of Bray-Curtis values calculated from filtered rarefied shotgun data. (**B**) The groups were largely split based on their differences in the proportion of Gram-positive and Gram-negative phyla in shotgun datasets was plotted for each group (chimpanzee and modern human, n=1; Neandertals, n=3). Error bars represent standard deviation.

Extended Data Figure 10: (**A**) Reads rom El Sidron 2 were mapped onto shared *Neisseria* genes (*i.e.* those gene regions shared between all of the species), and the resulting DNA fragments were aligned in MUGSY and compared with RAxML and bootstrapped with 100 iterations. (**B**) Phylogenetic analysis of whooping cough in Neandertals. Shared genomic regions within publically available *Bordetella* genomes were compared to ancient *Bordetella* reads from El Sidron Neandertals using RAxML with 1,000 iterations (bootstrap values).

Figure 1

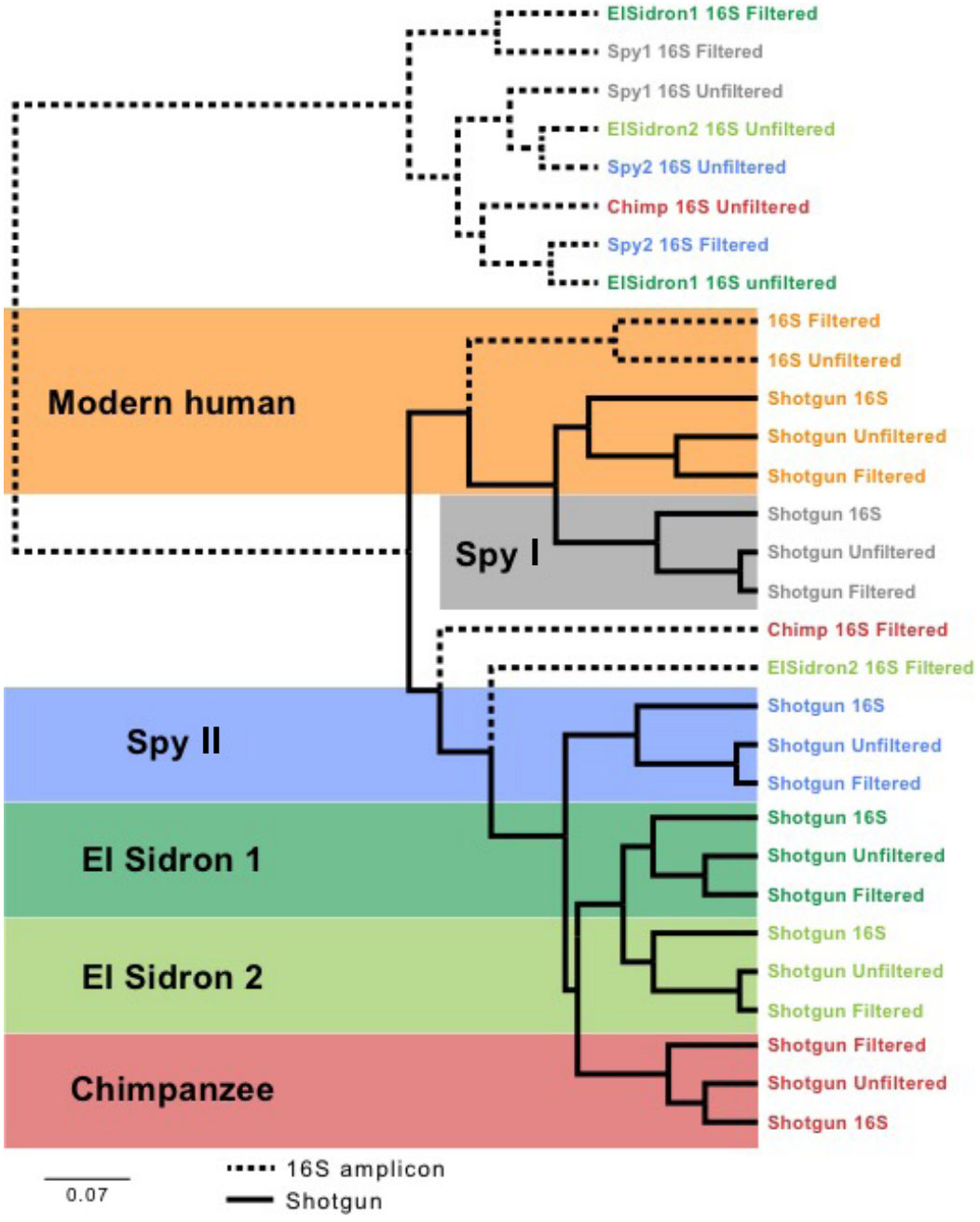
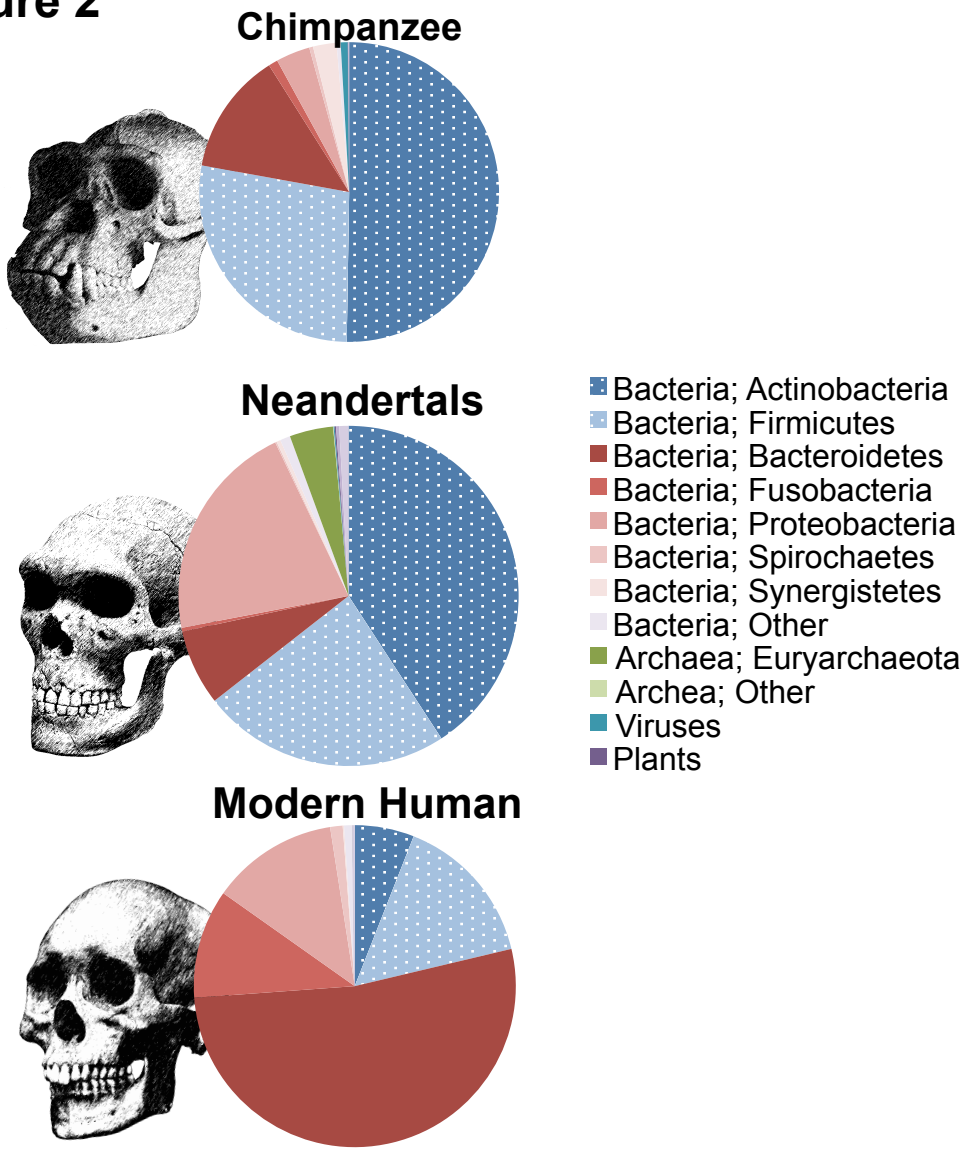

ElSidron1 16S Filtered
Spy1 16S Filtered
Spy1 16S Unfiltered
ElSidron2 16S Unfiltered
Spy2 16S Unfiltered
Chimp 16S Unfiltered
Spy2 16S Filtered
ElSidron1 16S unfiltered
Modern human
16S Filtered
16S Unfiltered
Shotgun 16S
Shotgun Unfiltered
Shotgun Filtered
Spy I
Shotgun 16S
Shotgun Unfiltered
Shotgun Filtered
Chimp 16S Filtered
ElSidron2 16S Filtered
Spy II
Shotgun 16S
Shotgun Unfiltered
Shotgun Filtered
El Sidron 1
Shotgun 16S
Shotgun Unfiltered
Shotgun Filtered
El Sidron 2
Shotgun 16S
Shotgun Unfiltered
Shotgun Filtered
Chimpanzee
Shotgun Filtered
Shotgun Unfiltered
Shotgun 16S
0.07
16S amplicon
Shotgun

Figure 2

A


Chimpanzee
Neandertals
Modern Human
Bacteria; Actinobacteria
Bacteria; Firmicutes
Bacteria; Bacteroidetes
Bacteria; Fusobacteria
Bacteria; Proteobacteria
Bacteria; Spirochaetes
Bacteria; Synergistetes
Bacteria; Other
Archaea; Euryarchaeota
Archea; Other
Viruses
Plants

B

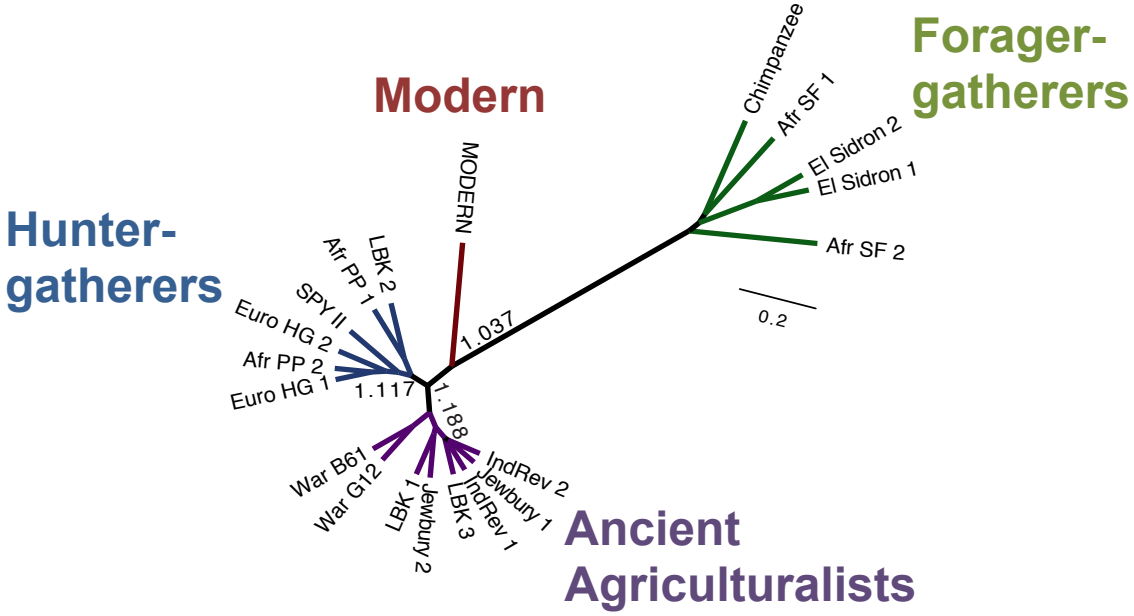
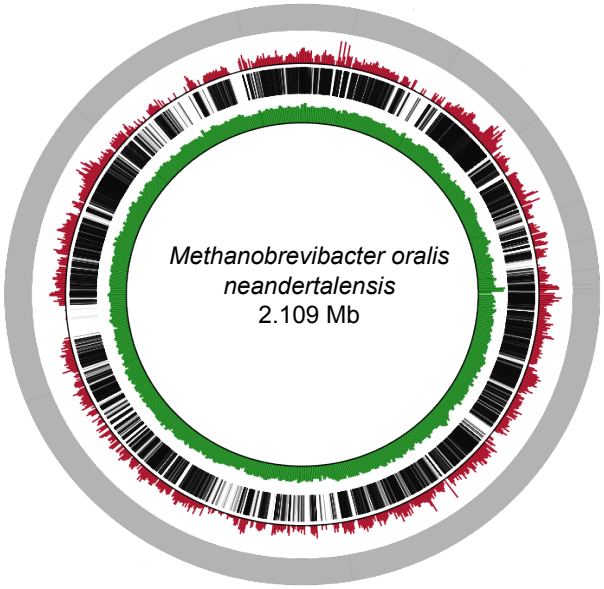
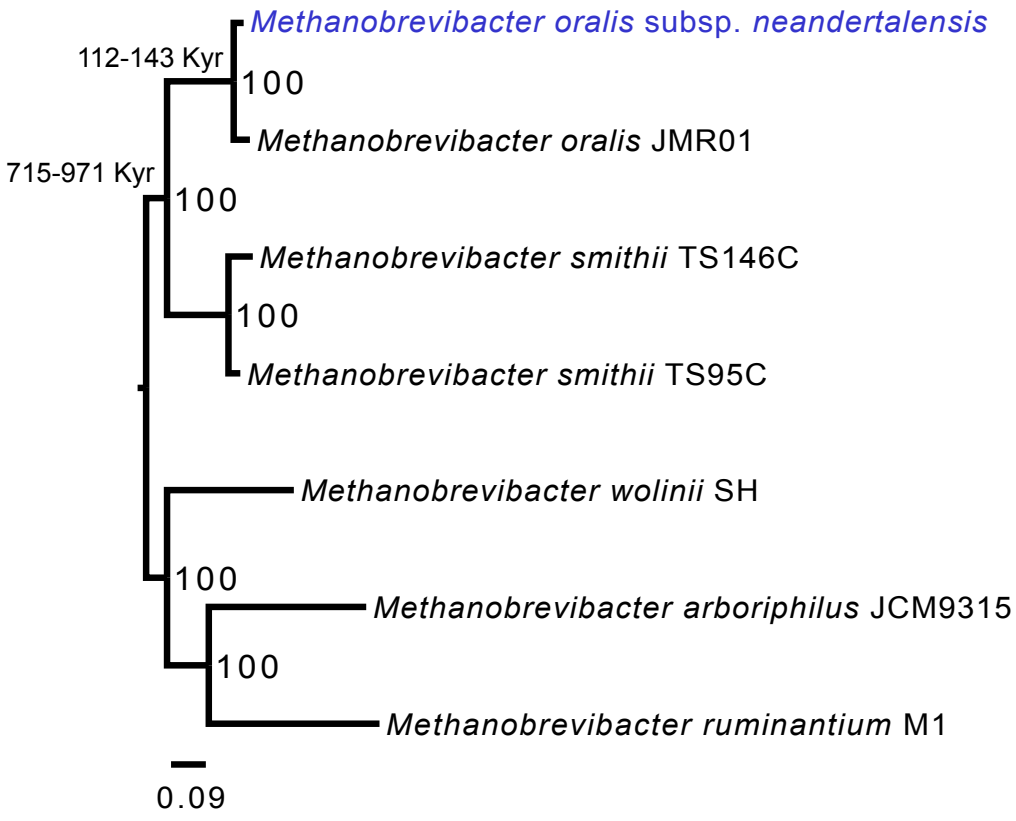

Modern
Forager-gatherers
Hunter-gatherers
Ancient Agriculturalists
MODERN
Chimpanzee
Afr SF 1
El Sidron 2
El Sidron 1
Afr SF 2
0.2
1.037
1.117
1.188
LBK 2
Afr PP 1
SPY II
Euro HG 2
Afr PP 2
Euro HG 1
War B61
War G12
LBK 1
Jewbury 2
LBK 3
IndRev 1
Jewbury 1
IndRev 2

Figure 3

A

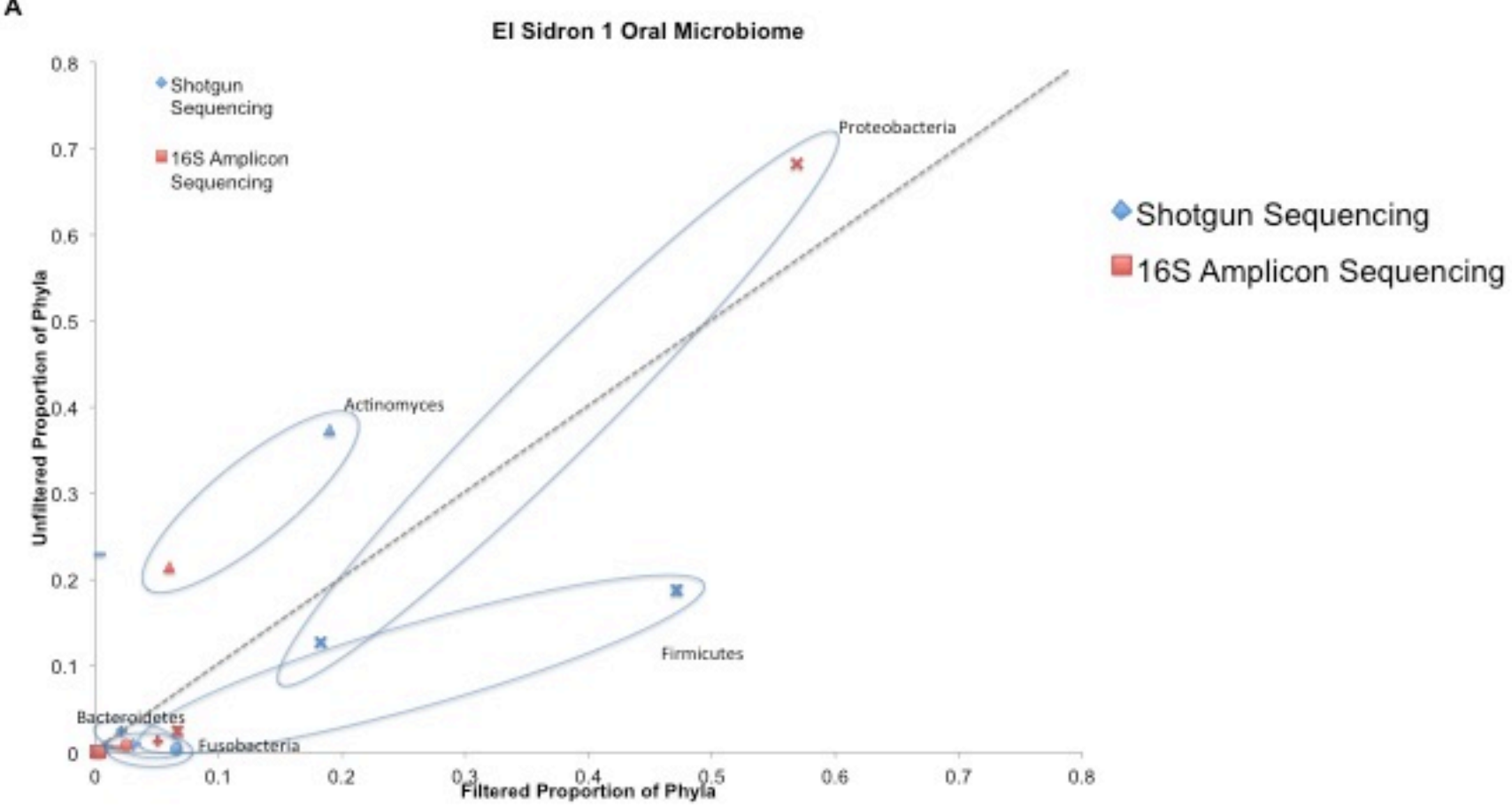

Methanobrevibacter oralis
neandertalensis
2.109 Mb

B

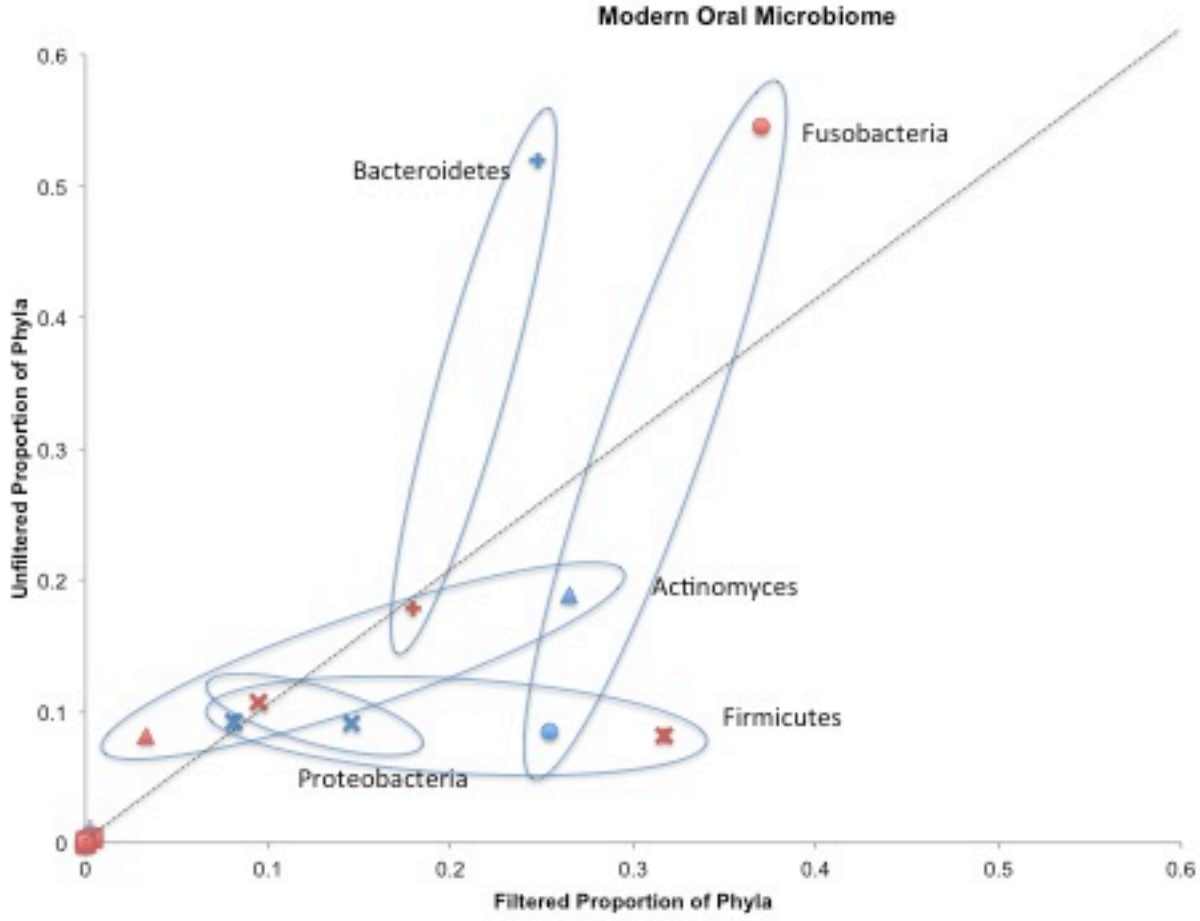

Methanobrevibacter oralis subsp. neandertalensis
112-143 Kyr
100
Methanobrevibacter oralis JMR01
715-971 Kyr
100
Methanobrevibacter smithii TS146C
100
Methanobrevibacter smithii TS95C
Methanobrevibacter wolinii SH
100
Methanobrevibacter arboriphilus JCM9315
100
Methanobrevibacter ruminantium M1
0.09

Extended Data Figure 1

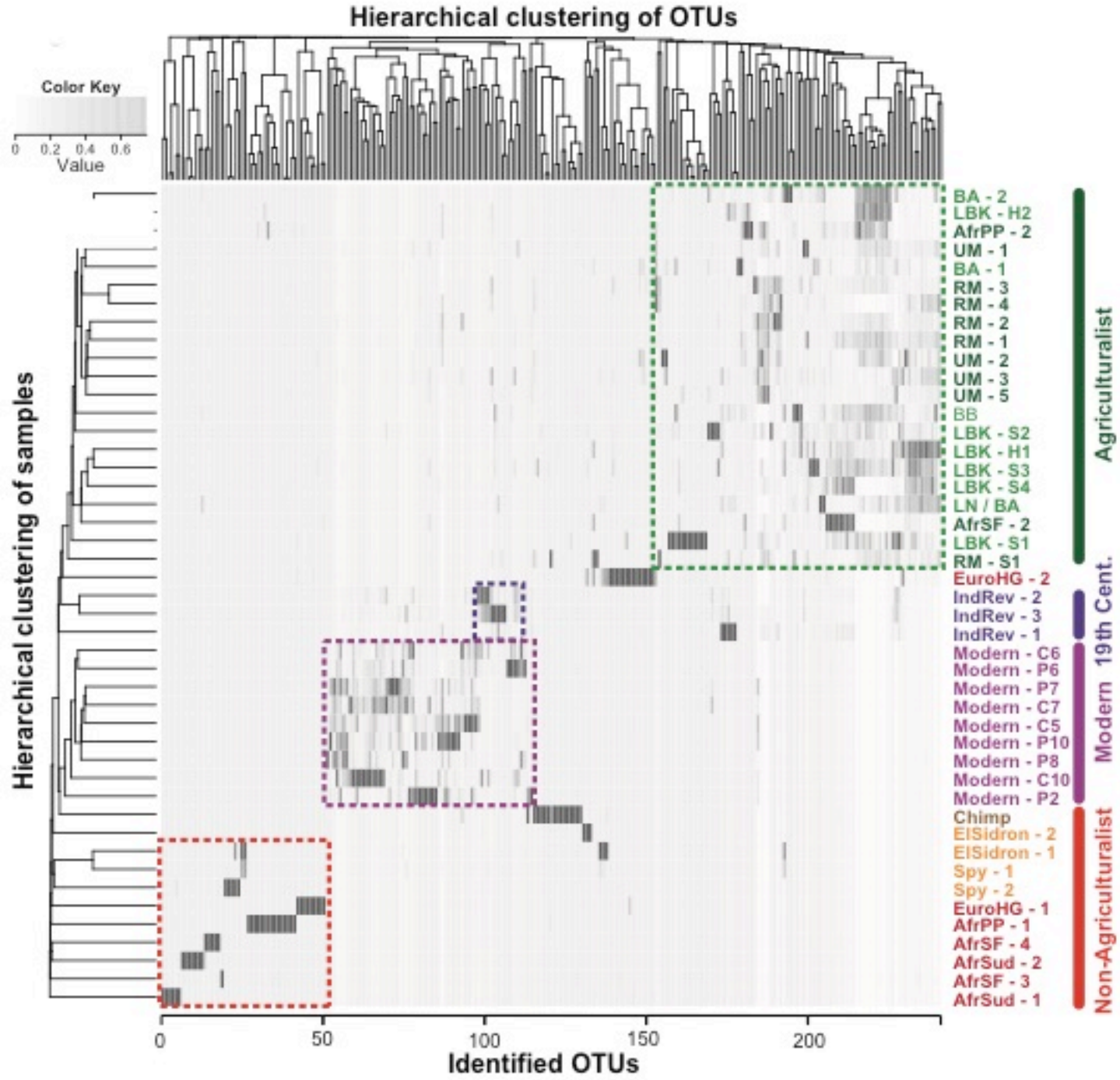
A



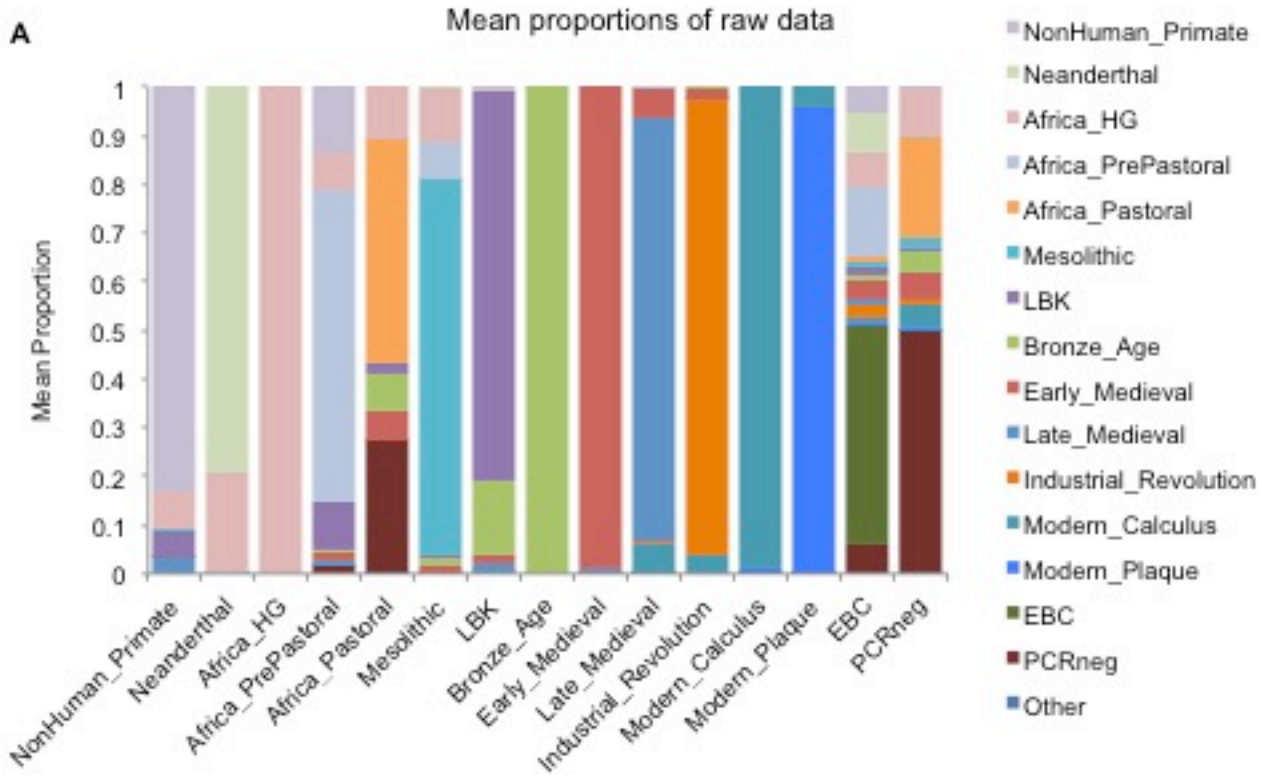
B

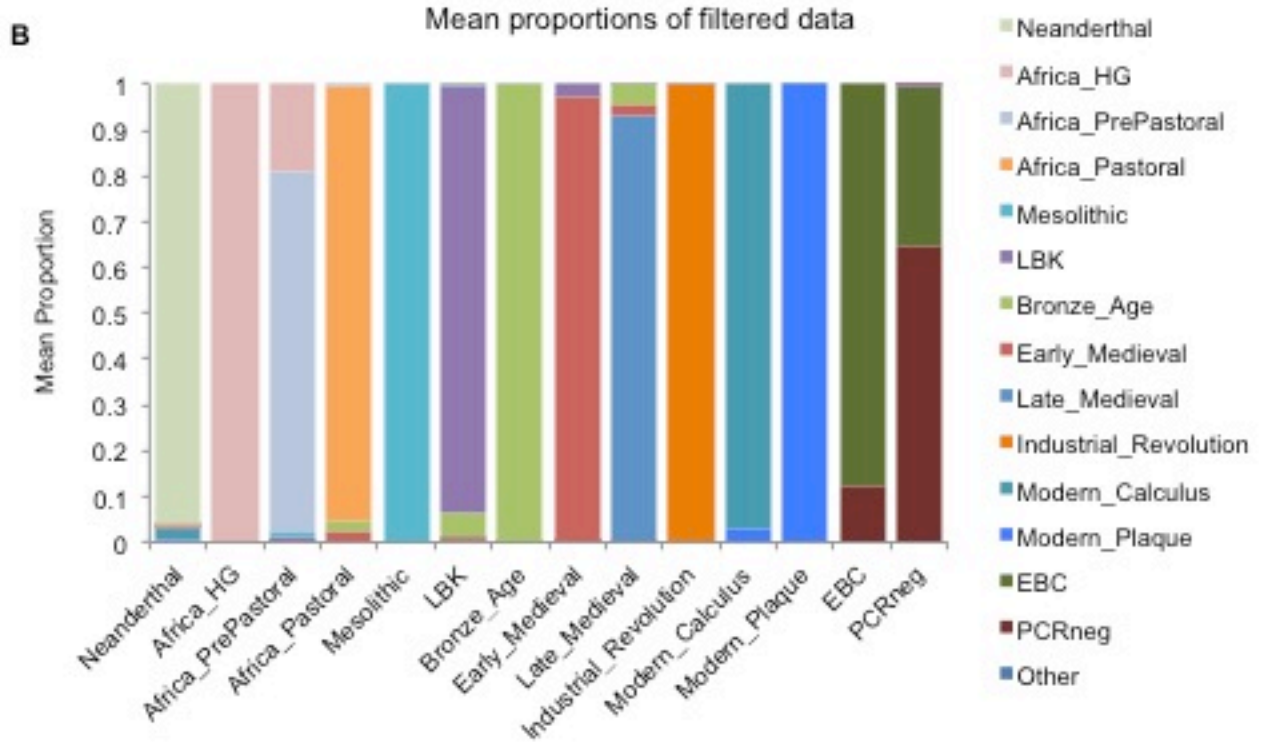


Extended Data Figure 2

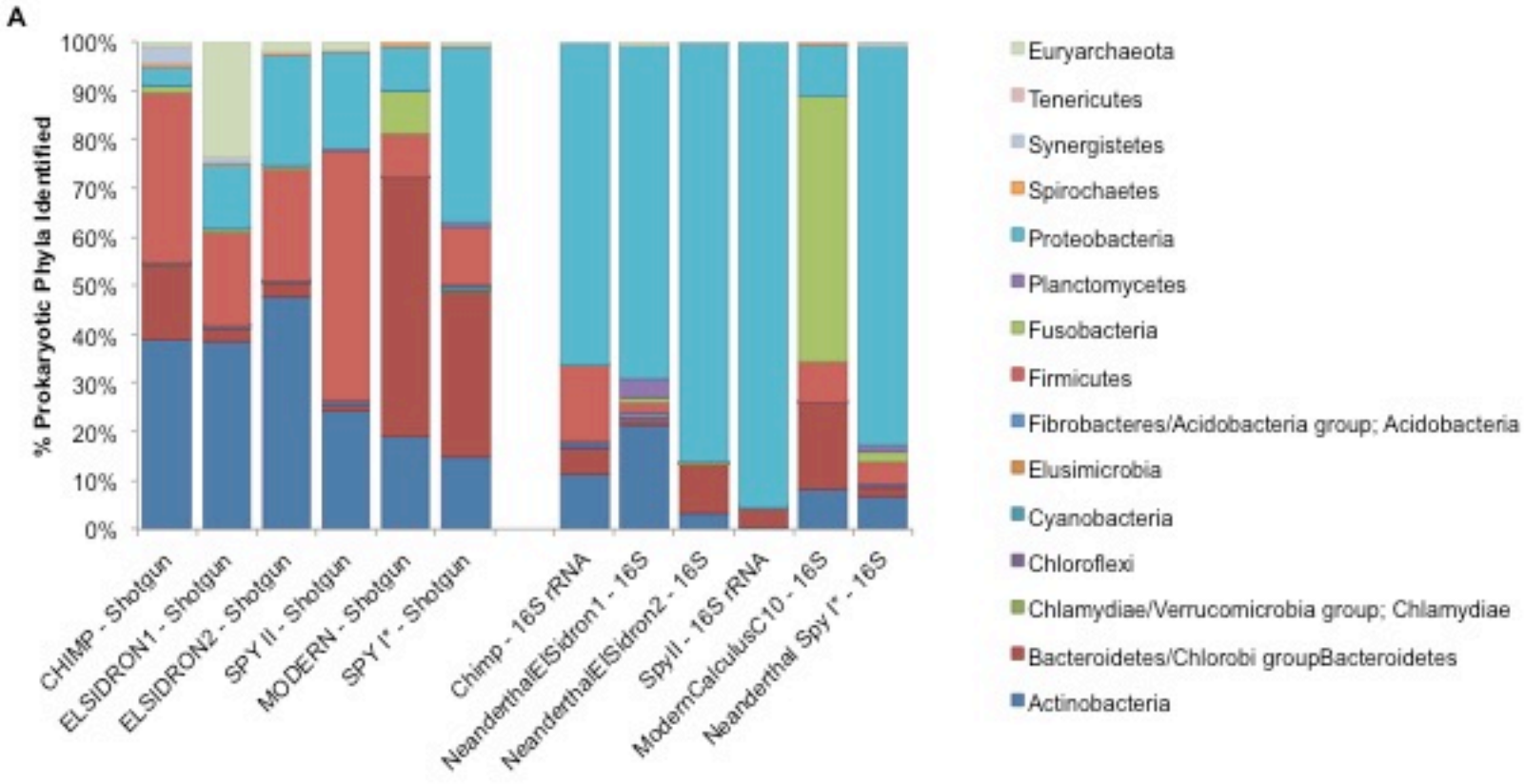


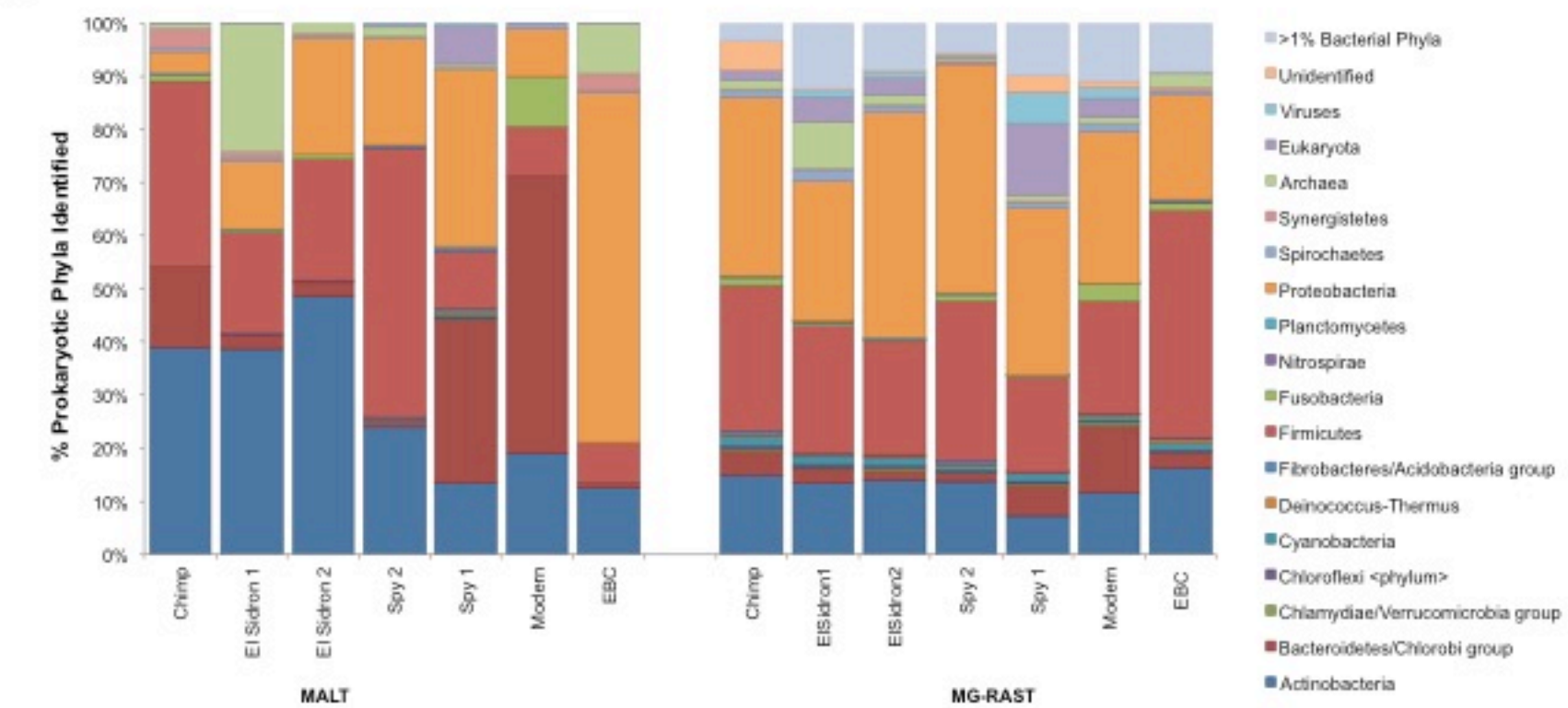
Extended Data Figure 3





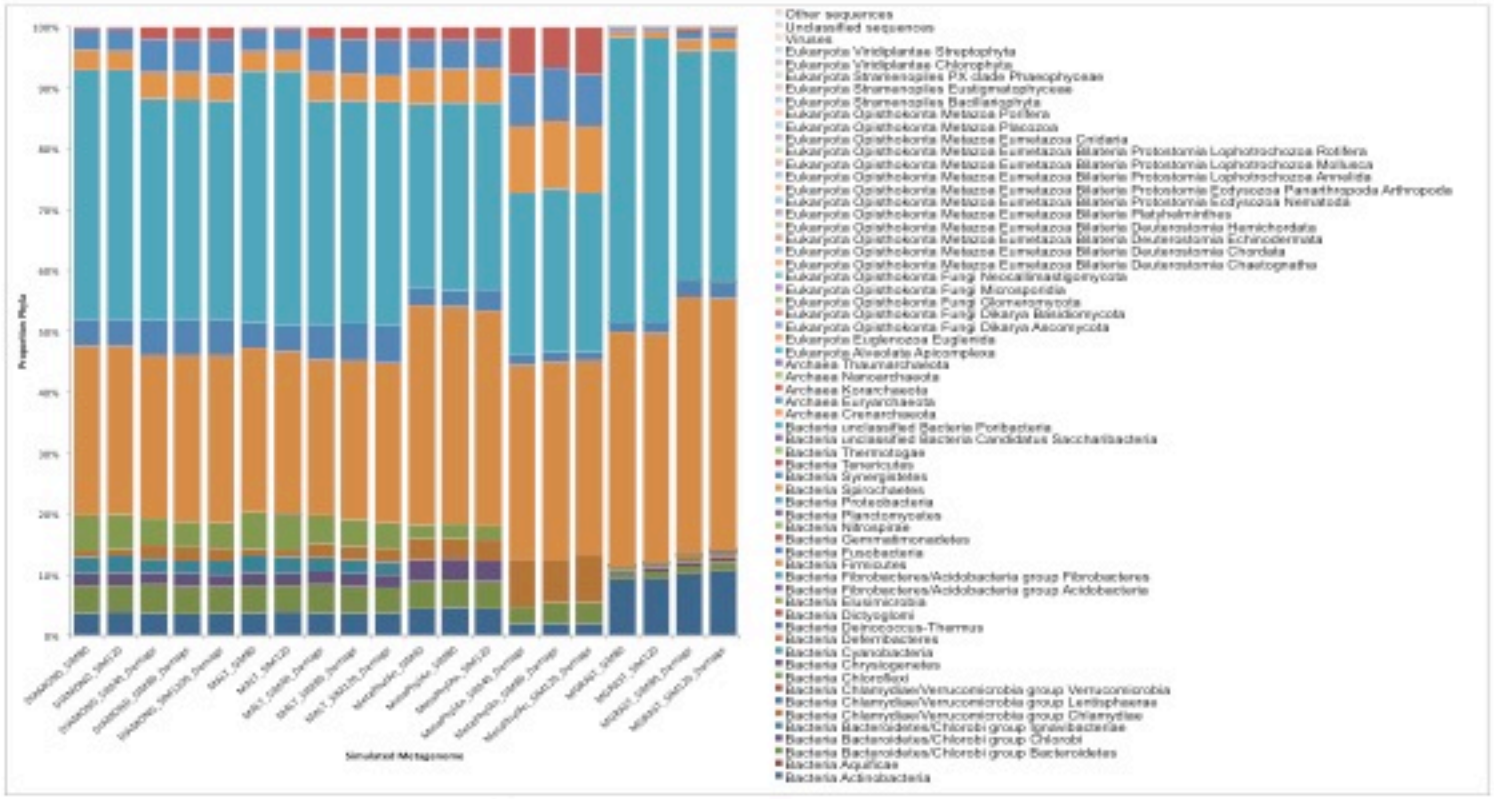
Extended Data Figure 4



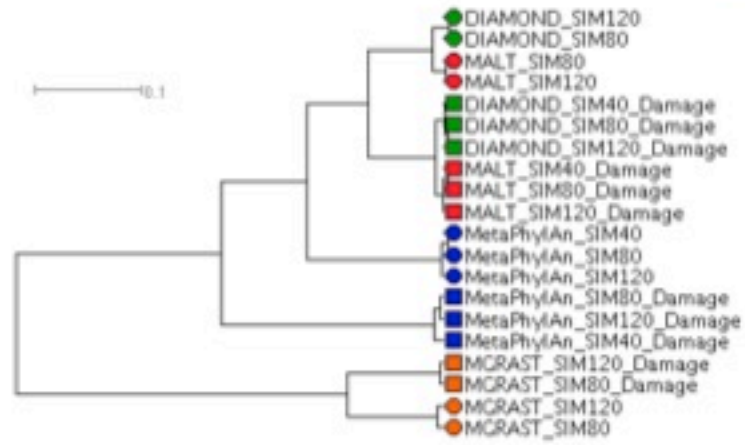


Extended Data Figure 5

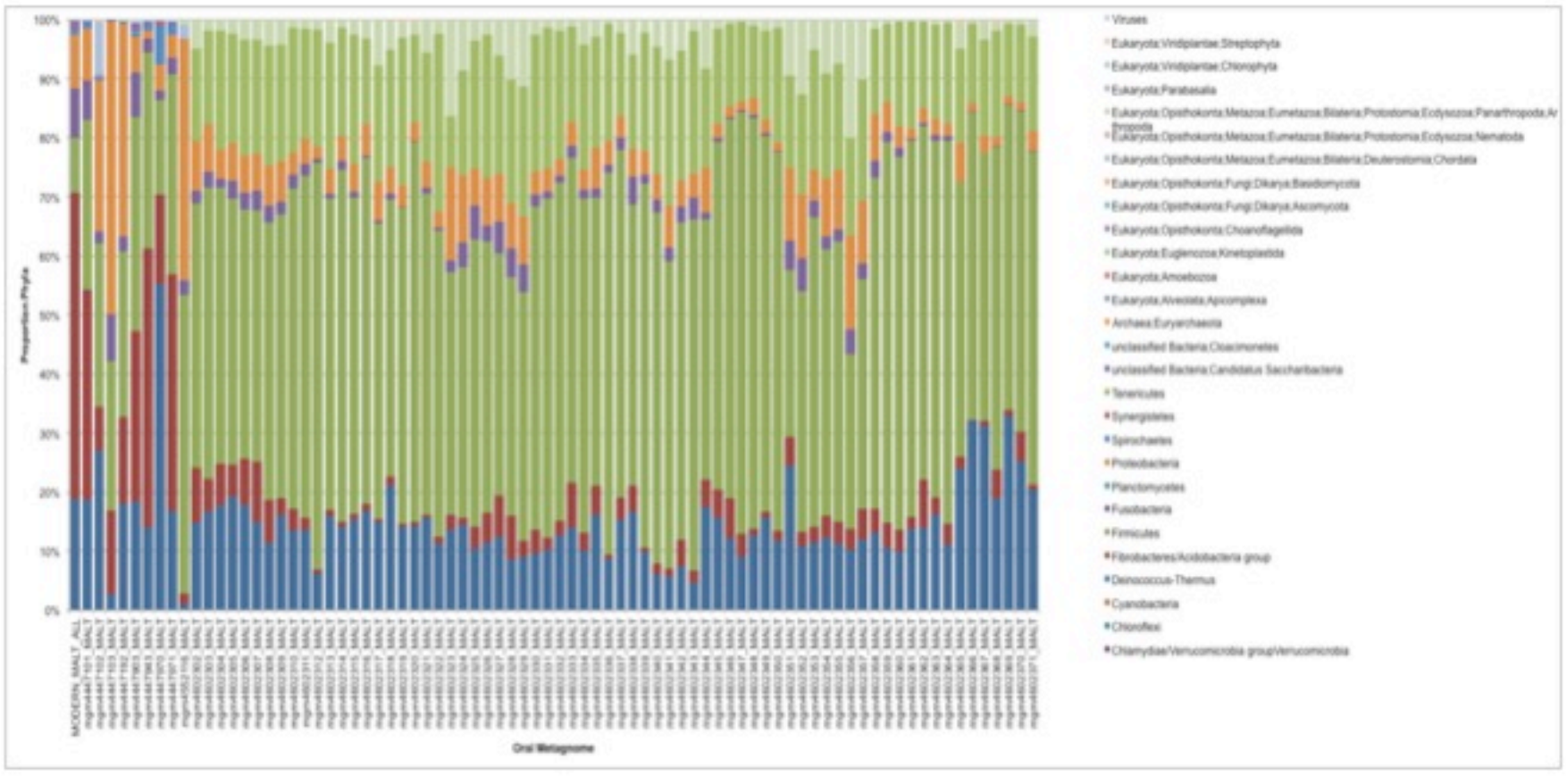
A



B

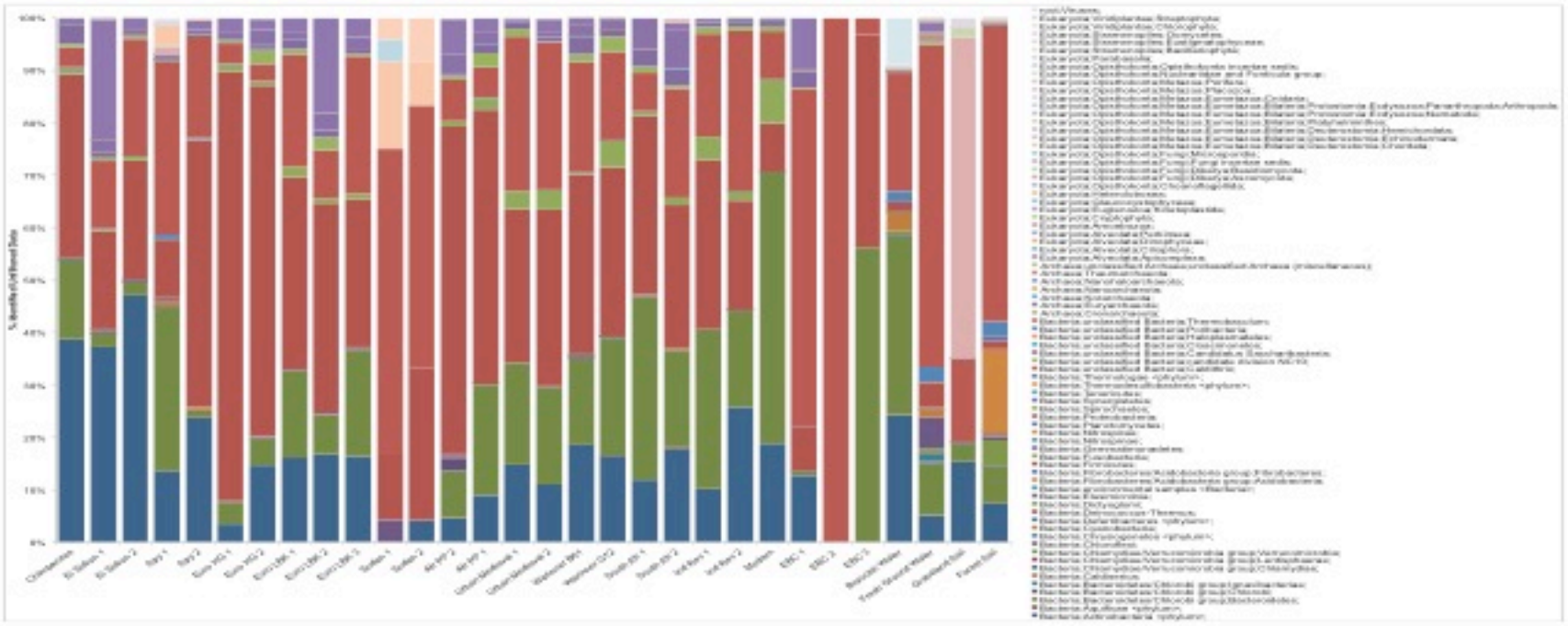


C

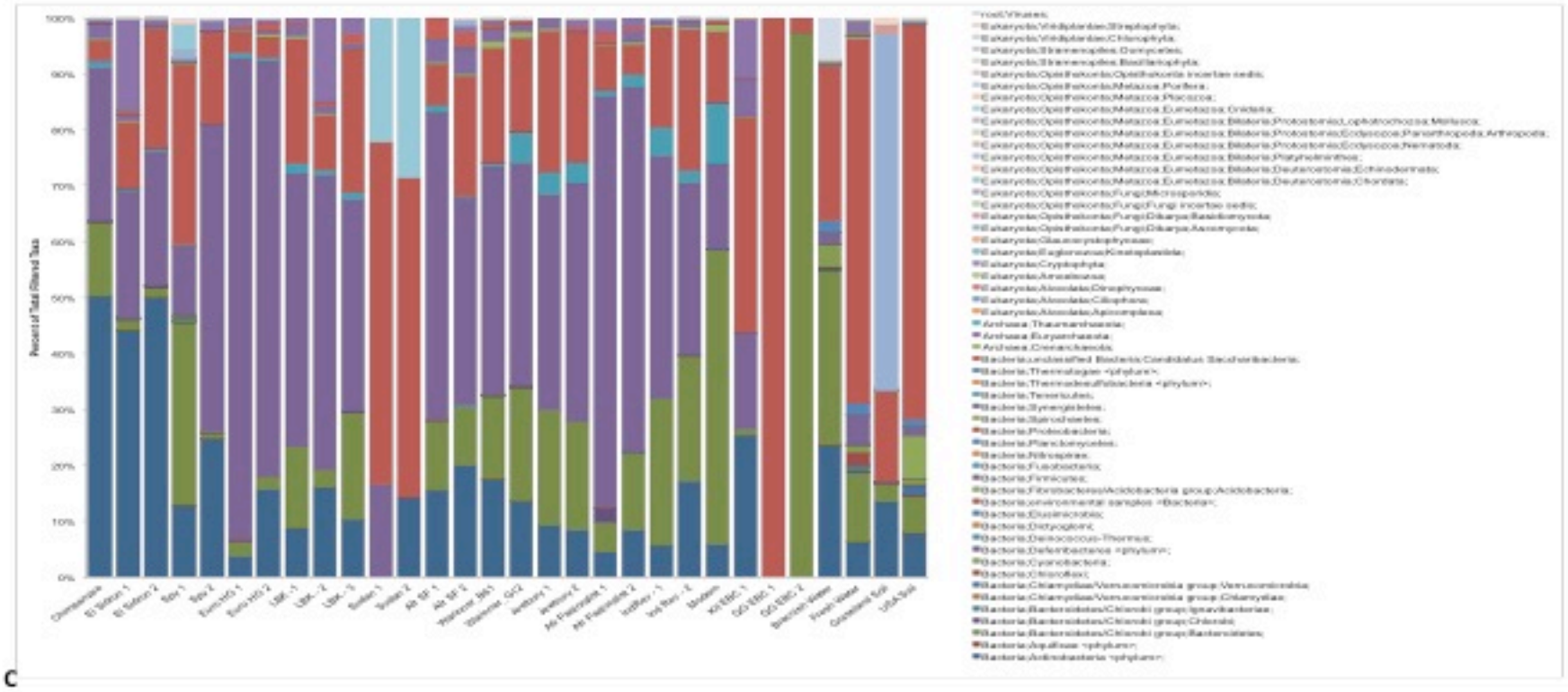


Extended Data Figure 6

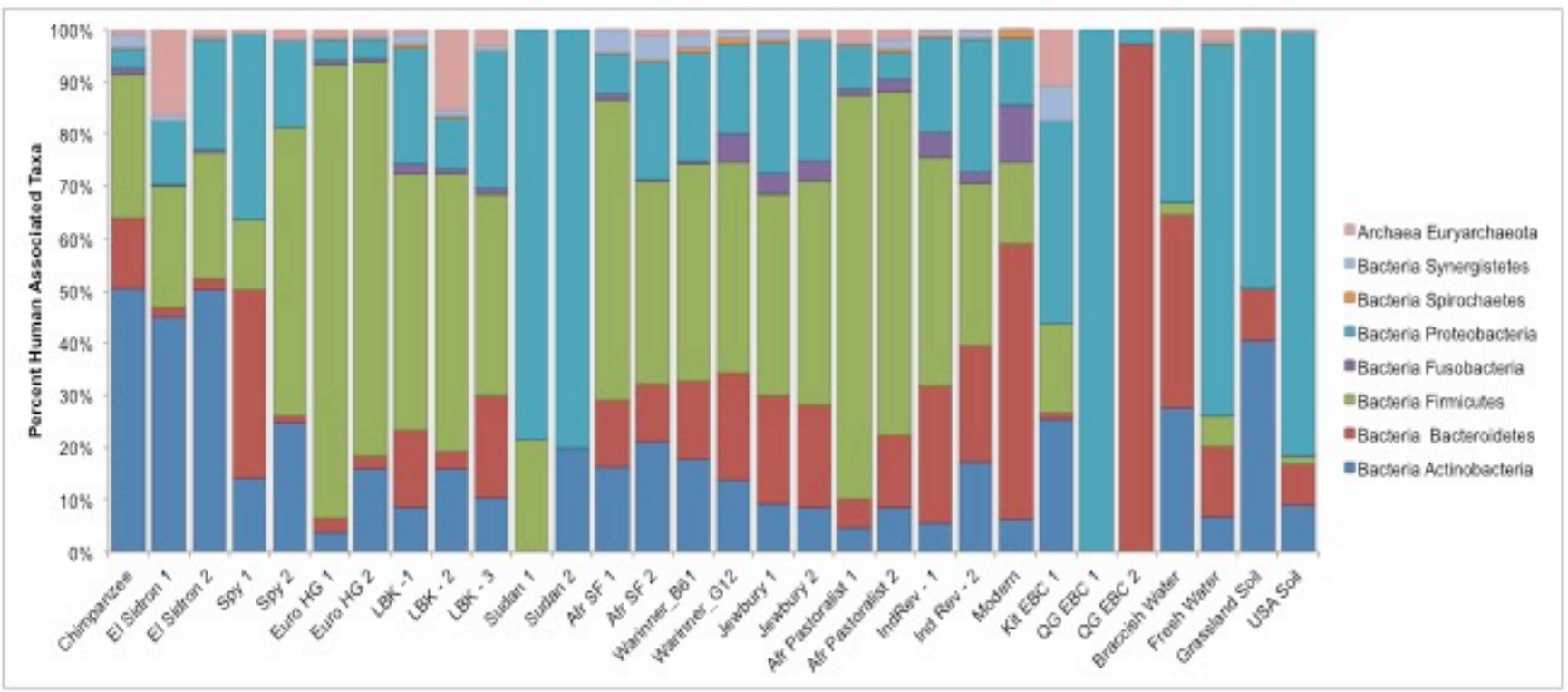
A



B

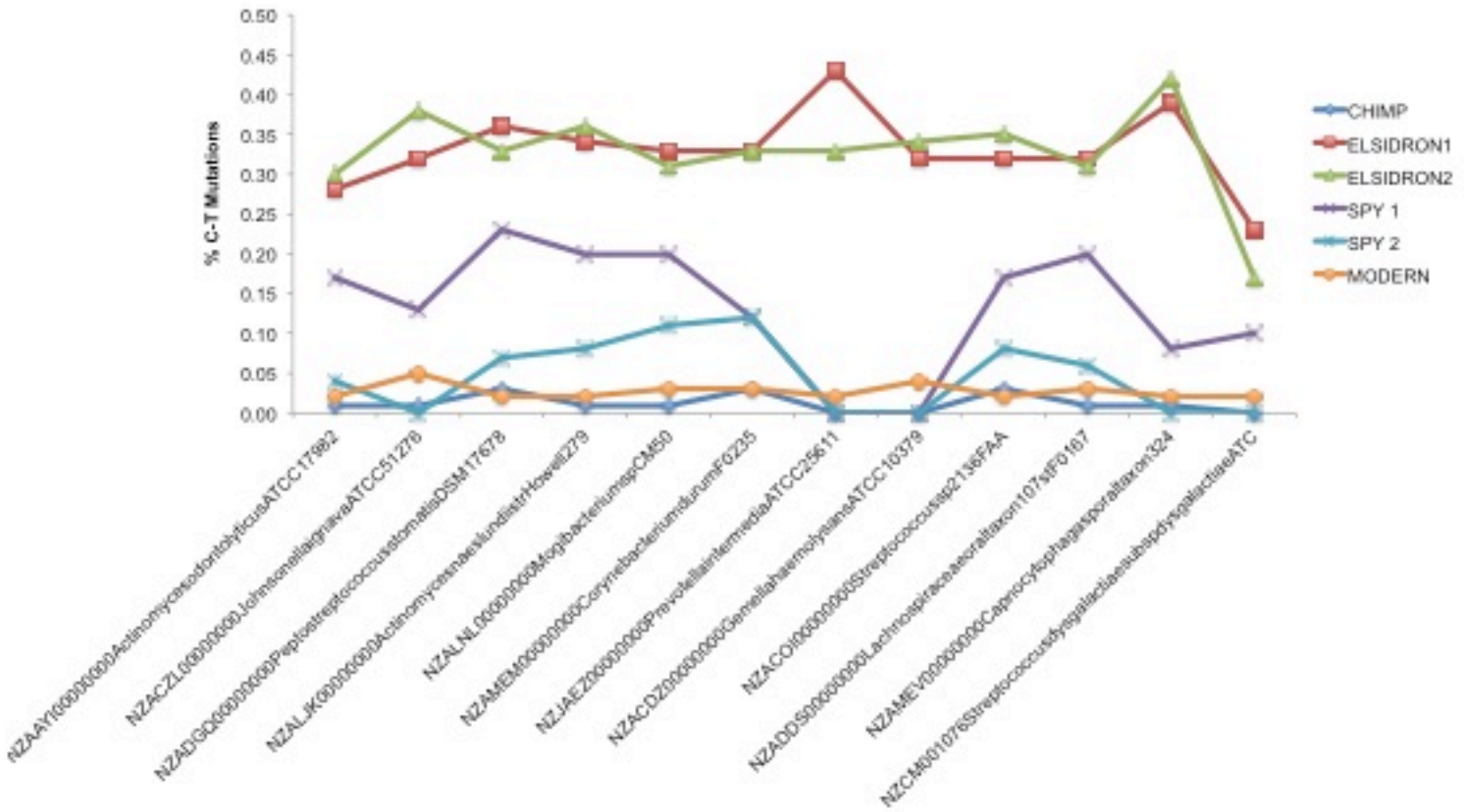


C

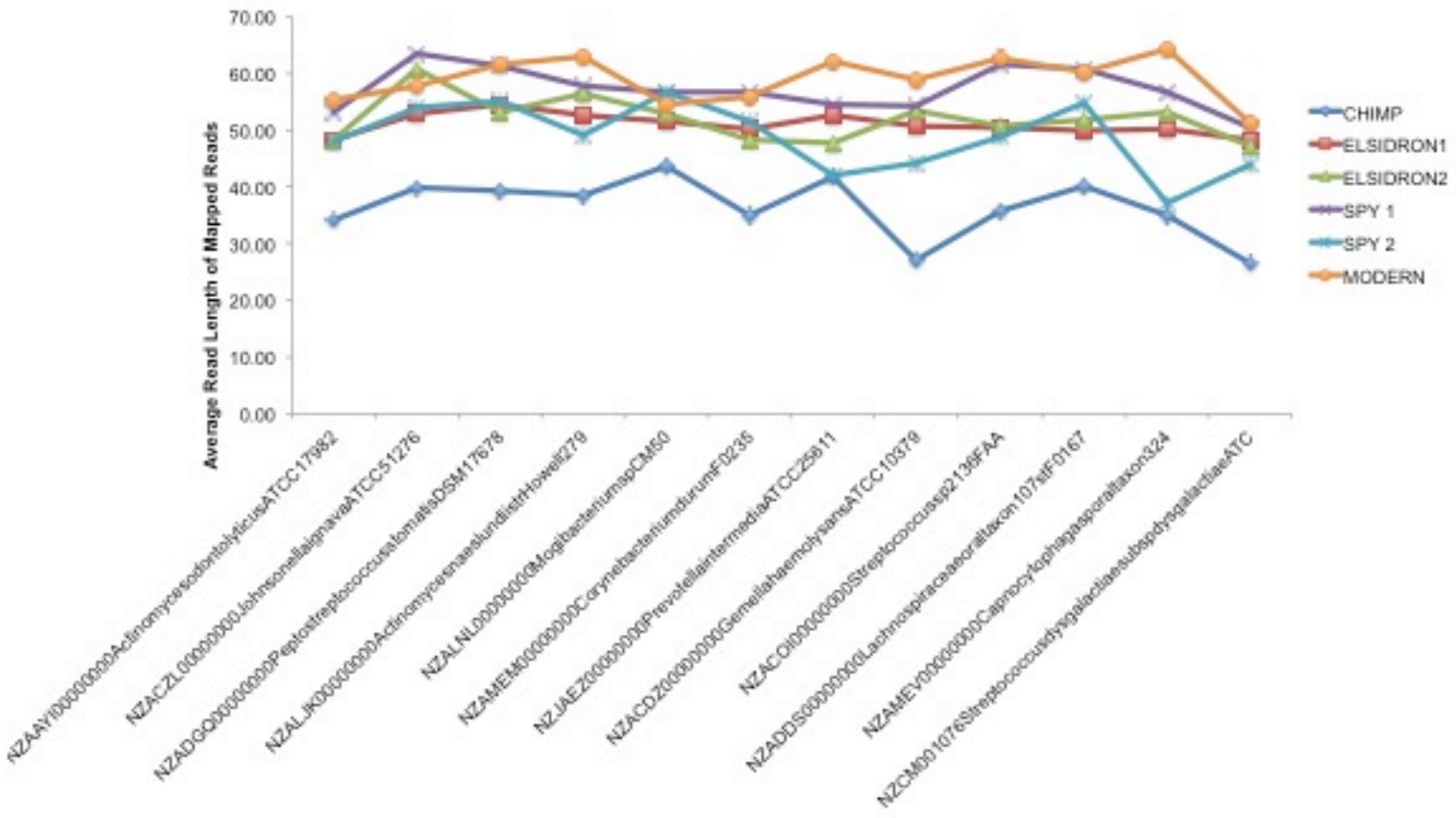


Extended Data Figure 7

A

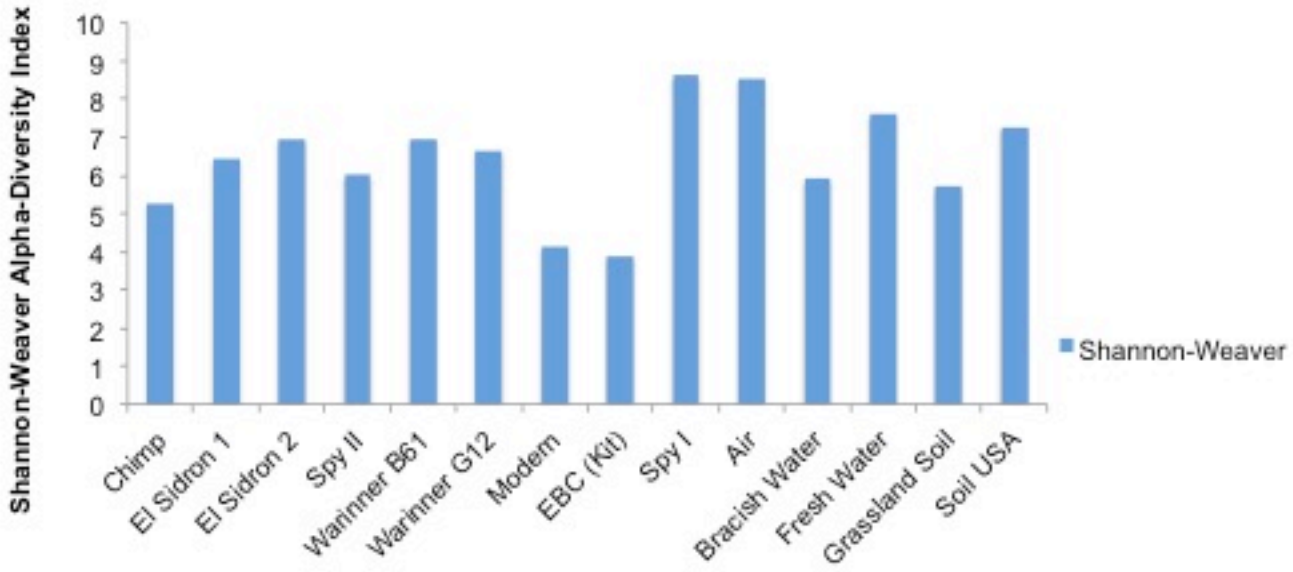


B

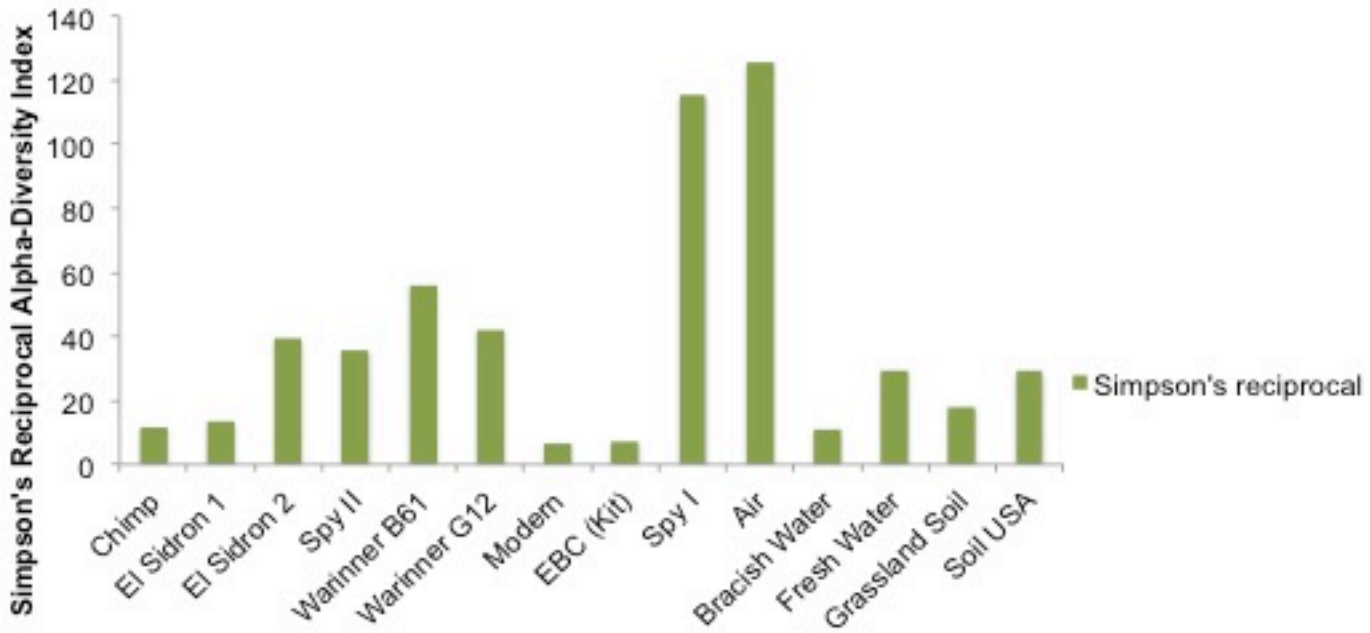


Extended Data Figure 8

A

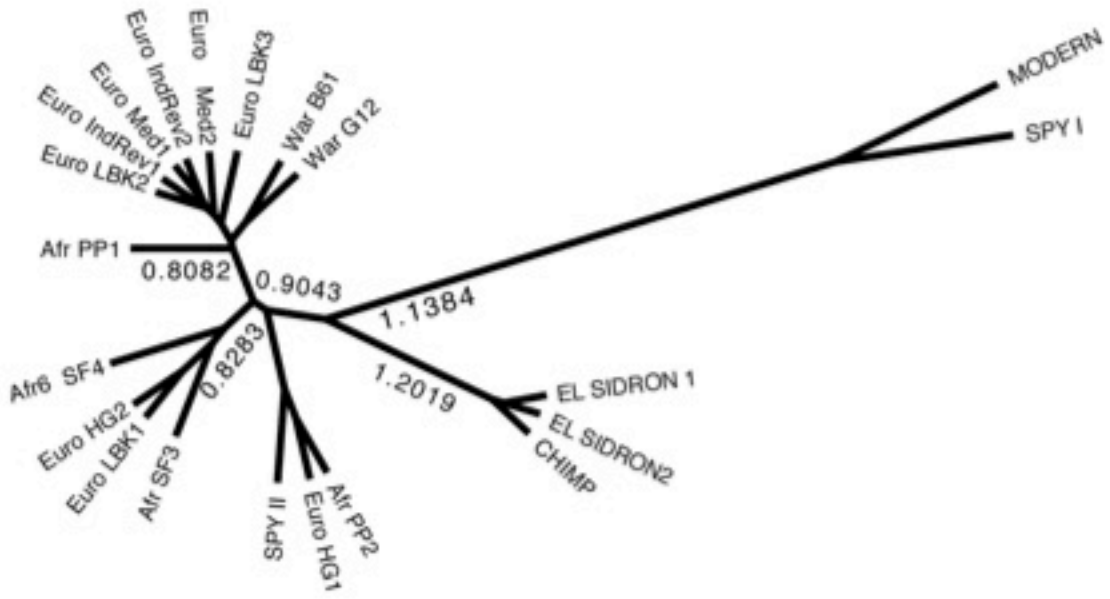


B

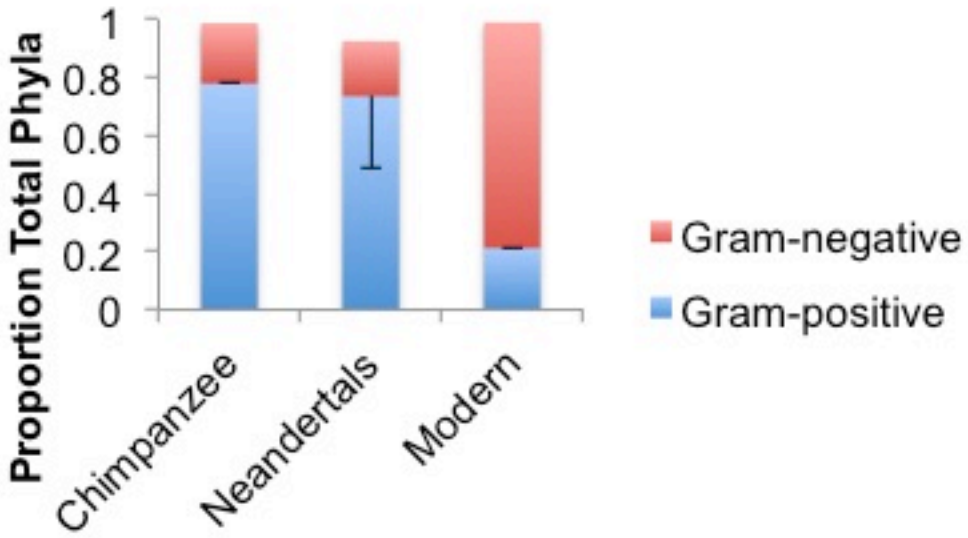


Extended Data Figure 9

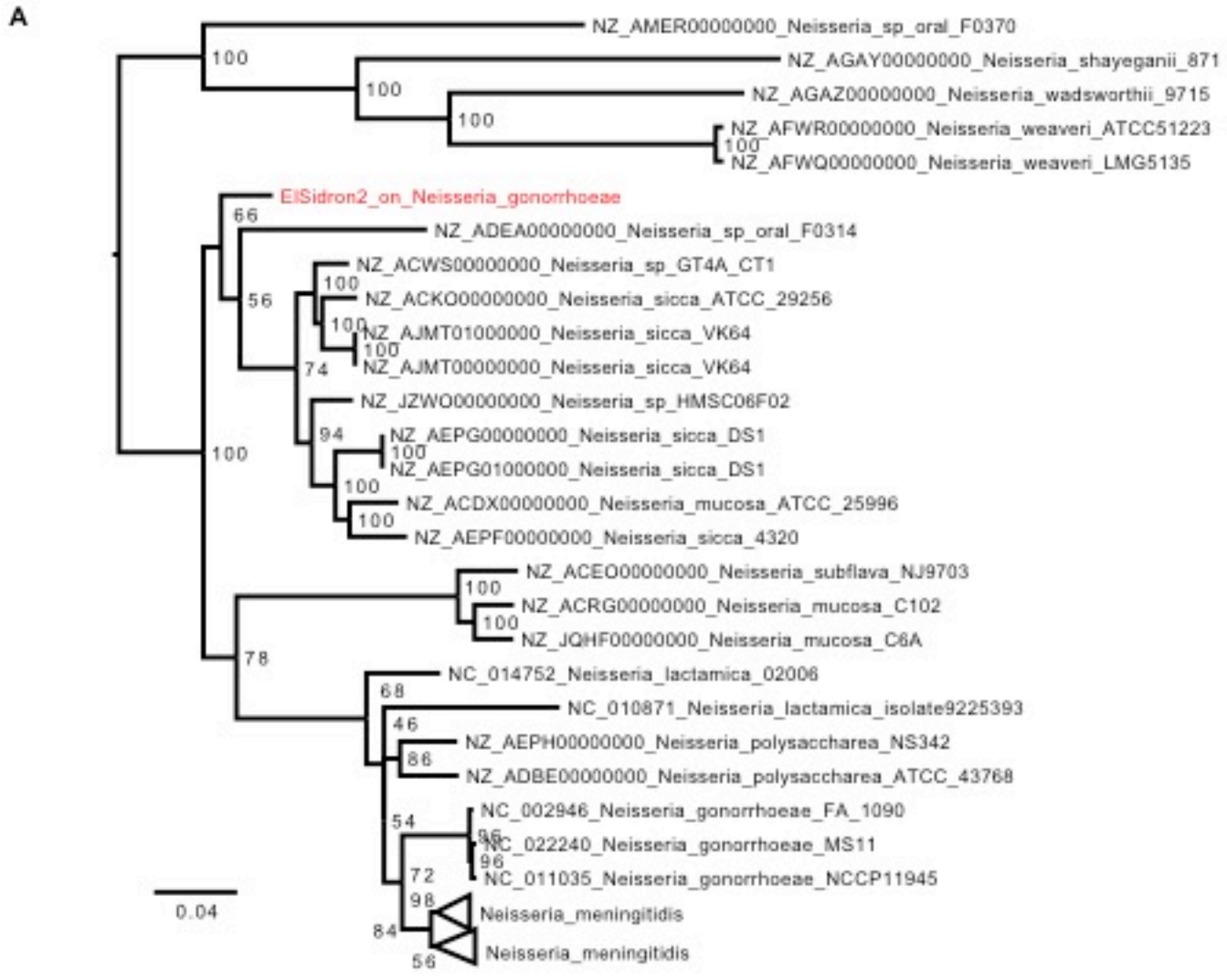
A



B



Extended Data Figure 10



Supplemental Materials – Table of Contents

Section I: Archaeological Context and Sample Descriptions

A total of 55 samples were analysed within this study, including Neandertals (n=5), a great ape (n=1), monkeys (n=2), ancient Europeans (n=30), ancient Africans (n=8), and present-day humans (n= 9). The Australian Centre for Ancient DNA (ACAD) and museum collection sample numbers, sample information, dating results, and classifications based on culture, group and period are given in Table S1. Samples were classified according to both geographic and cultural information (Culture or Group; Table S1) or chronographic information (Period; Table S1). The archaeological context of several European dental calculus samples and their cultures (*i.e.* Jewbury, Raunds, St. Helen on Walls, Bronze Age, and LBK Halberstadt) have been described previously by Adler *et al*[1]. Similarly, the context for new samples from a Linear Band Pottery culture (LBK) site near Stuttgart-Muhlhausen, Germany has also been previously described[2]. In addition, modern samples from healthy, 22-40 year old male and female volunteers at the University of Adelaide, Australia were included from a previous study[1]. The archaeological and anthropological context surrounding the remaining ancient and historic dental calculus samples is described within the following section.

Neandertals

The Neandertal dental calculus examined within this study was accessed from several different collections: Royal Belgian Institute of Natural Sciences in Brussels, Belgium, the Museo Nacional de Ciencias Naturales, in Madrid, Spain, and University of Florence in Florence, Italy.

The two Neandertal specimens from Spy cave, Belgium were obtained from separate Neandertal specimens, Spy I and Spy II, from the Royal Belgian Institute of

Natural Sciences in Brussels, Belgium, and were dated in 2009 to be ~36,000 yBP[3]. The Spy I supra-gingival calculus sample was collected from the lower M3 (Spy 94a), while the Spy II supra-gingival calculus sample was obtained from the lower M2 (92b). Although both individuals had calculus, their general oral health was excellent, with no cavities, abscesses, or major signs of periodontal disease. Spy I possesses tooth picking grooves on the upper and lower premolars and molars, while Spy II does not display any traces of tooth picking. This could be related to individual behaviour or to differential periodontal health. No DNA studies resolving the sex of these two specimens have been published, but skeletal evidence suggests that there is one male and one female specimen.

The dental calculus sample from El Sidrón 1 (1427c) was supra-gingival calculus from the lower molar of adult 2 (a young male, as determined by genetics and morphology of the mandible), whereas the calculus sample from El Sidrón 2 (1604) corresponds to adult 4 (female, as determined by skeletal morphology and canines)[4]. These two individuals have different mitochondrial lineages[4]. Although these two individuals are not maternally related, geological, archaeological, and genetic evidence has revealed that the El Sidrón individuals correspond to a contemporaneous social group of Neandertals, rapidly accumulated in the side gallery of a phreatic cave system through a collapse infill process. Analysis of the dentition in the young adult Neandertal male (El Sidrón 1) suggests he may have suffered from oral disease, as demonstrated by large calculus deposits and the presence of an abscess in the lower jaw[4]. Recent evidence also suggests that this specimen possessed a retained deciduous canine, which may have contributed to the levels of calculus present within this individual[5]. The location of these remains is currently the Museo

87 Nacional de Ciencias Naturales de Madrid, although the final destination will be the
88 Museo Arqueologico de Asturias, in Oviedo.
89 The dental calculus sample obtained from a Neandertal found in Breuil Grotta
90 in Italy has been dated at 36,600 yBP and has been described previously[6]. The
91 supragingival sample was removed from the lower left M2 occlusal tooth at the
92 University of Florence in Florence, Italy, although the molar is currently stored at the
93 Istituto Italiano di Paleontologia Umana Roma.
94
95 ***Non-human Primates***
96 Calculus from three different adult non-human primates, including a
97 chimpanzee, macaque, and baboon, were obtained from the Odontological Collection
98 at The Royal College of Surgeons, London. The chimpanzee calculus sample (ACAD
99 12873) was obtained from an individual that had been shot in the wild in Sierra Leone
100 in 1948. The macaque and baboon samples are from unknown provenance. The
101 macaque and baboon specimens were unique within the museum, so it is likely they
102 came from personal collections or were even living in captivity within England at the
103 time of death. None of the non-human primates had signs of oral disease or caries,
104 although the size of calculus between teeth and between individuals varied wildly.
105
106 ***Sudanese skeletons***
107 Dental calculus was removed from two skeletons excavated at Al Khiday 2, a
108 site located along the White Nile in the Sudan (16D4, grave 93 and 96). Both
109 specimens are female and full length descriptions have been provided previously,
110 although these samples do not have radiocarbon dates[7]. Briefly, the skeletons are
111 Neolithic (mid 5th Millenium BC) and form part of an assemblage of 190 individuals

112 excavated from a multi-period cemetery. The Neolithic skeletons were found in
113 shallow circular pits. At this site, pots and ornamental objects were found in some
114 graves, which were culturally associated to the El Shaheinab phase[8]. The Neolithic
115 groups along the Nile Valley are thought to be pastoralists, but there is little
116 information about their diet, as no well-preserved settlements have been found.
117 Faunal remains of domestic and wild animals[9] were accompanied by both wild and
118 domesticated cereals, such as wheat and barley[10], and were found in funerary
119 contexts, suggesting a mixed diet. This diet likely supported the extensive calculus
120 formation identified on these two Sudanese skeletons[7].
121
122 ***South African skeletons***
123 The eight individuals from South Africa included within this study were taken
124 from a series of over 40 Later Stone Age (LSA) skeletons stored in the Department of
125 Human Biology of the University of Cape Town[11]. All skeletons are from the Western
126 Cape Province, and the only selection factors were a radiocarbon date older than 350
127 yBP and the presence of significant calculus deposits on the teeth.
128 The eight individuals selected were divided into two groups based on the age
129 of their skeletons. South African samples from the Pastoralist Period (AfrPP) AfrPP1
130 (UCT 587), AfrPP2 (UCT 67), and AfrPP3 (UCT 157) are all less than 1,000 years
131 old, while the other group of African hunter-gatherers (AfrSF) included seven
132 individuals, who are all older than ~2,000 years (AfrSF). The oldest specimen was
133 AfrSF2 (UCT 373), who died just short of 4,000 years ago. All ten individuals had
134 advanced occlusal wear on the teeth, and all but AfrPP3 (UC 157) had heavy or
135 extremely heavy wear on the incisors.
136

South African Specimens >2,000 yBP (AfrSF)

All of these individuals died more than 2,000 yBP, and are from coastal or adjacent areas ranging from the Cape Peninsula in the south to Elands Bay, about 200 km north of Cape Town, South Africa. All are from hunter-gatherer contexts, as pastoralists did not enter the region until 2,000 yBP[12,13]. The five individuals are female, and all exhibit advanced occlusal tooth wear, although the general data health pattern is better than the Pastoralist Period group. Only two individuals had initial caries in the enamel, and both instances were relatively minor. Furthermore, no teeth had been lost antemortem.

This is oral health pattern is consistent with an active foraging lifestyle with limited access to simple starches and free sugars[14], and where the teeth are cleaned mechanically by chewing of high fibrous content food. All of the hunter-gatherers within this collection that had dental calculus were women, potentially indicating a sex based difference in calculus formation, potentially explained by the female dental masticatory regime. In historic Kalahari hunter-gatherers, women continually sample nuts, berries, and roots as they gather food during the day[15], which might result in increased rates of dental calculus.

South African Pastoralist Period Samples (AfrPP)

These individuals all died roughly between 800 and 600 yBP and lived on the south coast of the province or in the neighbouring hinterland. AfrPP1 (UCT 582) was a woman from a pastoralist (as opposed to a hunter-gatherer) group, but would still be considered from the LSA in terms of technology[16]. This female individual had one carious premolar, had lost both lower first molars antemortem, and had a severe abscess in the region of the mandibular third molar. Next, AfrPP2 (UCT 67) was a

male, but there is no archaeological information to confirm the lifestyle of this individual. AfrPP2 (UCT 67a) also had extensive periodontitis and had lost five teeth from the left side of the maxilla before death. A third sample AfrPP3 (UCT 157) within the Pastoralist Period group was initially analysed, but did not pass bioinformatic filtering due to the low numbers of identified phyla. This specimen was also probably from a foraging community, although there is no archaeological information to confirm the lifestyle of this individual. AfrPP3 (UCT 157) had three carious teeth, one of which was extensively damaged with over half of the crown missing.

All of the Pastoralist Period African individuals showed relatively poor dental health. It is possible that agricultural foodstuffs were traded in from the Bantu-speaking groups to the east of the region who were settled in the area of the Transkei by 800 years ago[17]. The general pattern of tooth loss and decay is in stark contrast to that seen in regional hunter-gatherers. The number of teeth with caries accounts for 8% of teeth recovered for the three individuals, and this rises to ~20% if the antemortem losses are considered. This is well above the expected numbers seen in hunter-gatherers[18] and is more consistent with a diet consisting of a mix of gathered and agricultural foods. The relatively low fluorine levels on the south coast of the Cape Province complicate this interpretation, as people raised in the region are generally more susceptible to dental diseases[18].

Post-Industrial Revolution German Samples

Dental calculus samples were obtained from Hettstedt, located within the Mansfeld district in Saxony-Anhalt, Germany. The Hettstedt cemetery was set up in 1853 and excavated by the State Office for Heritage Management and Archaeology

187 Saxony-Anhalt and Heritage Museum, Halle, Germany in 2010. The 75 burials were
188 osteologically analyzed in an unpublished Master thesis by Klapdohr *et al.* in 2013.
189 Socioeconomically, the Hettstedtt population is characterized by the mining and
190 metallurgical work in the town. At the beginning of the 19th century, the inhabitants
191 died at an early age, and only 5-7 % of the population reached their 60th year. By the
192 end of the 19th century, life expectancy increased noticeably, likely due to improved
193 hygiene and dietary changes, as well as the replacement of gruel as the main dietary
194 component with a combination of cereals, potatoes, and milk. In addition, tea, cocoa,
195 sugar, and tobacco no longer were luxury goods available to the elite only, and sugar
196 consumption rose markedly from 2.4 kg per capita from the year 1840 to 4.7 kg in
197 1861 and 7.3 kg in 1875[19].
198 For 63 out of the 75 Individuals (84%), dental pathologies could be evaluated.
199 Tooth loss during life was considerably higher in women than in men (60.9% and
200 39.1%, respectively). Approximately 20% of the adults exhibited an intermediate to
201 severe degree of dental plaque formation. On average, each of the 63 individuals
202 exhibited 15 carious lesions. Caries frequency was 100%, and caries prevalence was
203 over 50%. Even children and juveniles exhibited a large proportion of teeth with
204 dental caries. Their poor dental health can probably be attributed to mushy and sticky
205 foodstuffs.

Section II: Additional Methods and Results

Sample Collection

Dental calculus samples were collected at the respective locations that housed the skeletal material (Table S1). First, the teeth were examined to identify the tooth with the largest calculus deposit. Once a ridge on the largest deposit was identified, the skeletal material was wrapped in aluminium foil, and the calculus was removed from a single tooth by applying pressure with a dental pick. The calculus from an individual tooth was then collected into the aluminium foil and poured into a labelled, non-breakable container for transport (*e.g.* a sterile plastic 2. mL screw cap tube or plastic bag). On average, each calculus sample was <0.01 gram and 2x3x2 mm in size. Photographs of representative calculus samples from each host species have been included for reference (Figure S1).

Sample preparation and DNA extraction

Upon arrival, all ancient samples were housed within a quarantine facility for ancient DNA research at The Australian Centre for Ancient DNA (ACAD) at The University of Adelaide, Australia. Before entry into the facility, sample containers were bleached and UV irradiated for 15 minutes to minimize exogenous microbial contamination. Samples were then stored at 4°C until DNA extraction was performed.

Prior to extraction, each sample was subjected to stringent decontamination procedures to reduce environmental contaminant DNA present on the outer surface of the dental calculus. The decontamination procedure is as follows. Each sample was individually placed in a sterile plastic dish and exposed to UV radiation for 15 minutes on each side. Next, the sample was submerged in 5% bleach and then purified water for 5 min, followed by immersion in 90% ethanol for 3 minutes to

231 remove any residual bleach. The sample was then air dried for 5 minutes within a
232 sterile container and then transferred to a sterile plastic bag. While inside the bag, the
233 calculus was pulverized with a steel hammer. The corner of the plastic bag was then
234 removed, and the sample was poured into a sterile 2 mL screw cap tube.
235 Once the powder was collected, DNA was immediately extracted as
236 previously described, with the following modifications[20]. Briefly, samples were de-
237 calcified by adding the powderized sample to a sterile tube containing 1.8 mL of 0.5
238 ethylenediaminetetraacetic acid (EDTA), 100 µL of 10% sodium dodecyl sulphate
239 (SDS), and 20 µL of 20 mg/mL proteinase K and left to rotate at 55°C for 20 hours.
240 Modern samples were extracted using the same protocol in a modern DNA lab at the
241 University of Adelaide, but the resulting DNA (from modern samples only) was
242 sheared according to manufacturer's recommendations to ancient DNA-like sizes
243 (~150 bp) via a Covaris sonication system prior to shotgun library preparations.
244 Extraction blank controls (EBCs) were included at the beginning and end of each
245 sample series (at approximately a ratio of one EBC per ten samples). Two empty
246 tubes were included as EBCs in each extraction; these tubes were treated as if it were
247 a powderized calculus sample. Released DNA was then bound to silica with 3 mL of
248 modified QG buffer (Qiagen) as previously described[20], pelleted, and washed twice in
249 80% ethanol. Next, the cleaned and dried silica was resuspended in 100 µL of Tris-
250 HCl buffer twice to elute the DNA. Eluted DNA was aliquoted and stored at -20°C
251 until amplification. Notably, all the DNA and RNA-free certified water (Invitrogen
252 Ultrapure distilled water) used to create the reagents was opened fresh or was frozen
253 upon aliquoting to prevent microbial growth and contamination.
254 DNA from calculus samples examined previously by Adler *et al.* was also
255 included in this study (Table S1)[1]; these samples were amplified according to the

protocol below. To compare extraction efficiencies, Late Medieval samples (ACAD 8812 and 8824) were extracted using both methods. The two different extraction methods produced bacterial OTUs that were not statistically different at the phyla level (Figure S2).

16S Ribosomal RNA Amplicon Library Construction and Analysis

Amplicon Library Preparation and DNA sequencing

The V4 region of the bacterial 16S ribosomal RNA (rRNA) encoding gene was targeted for amplification using degenerate Illumina fusion primers, as previously described[21]: forward primer 515F (AATGATACGGCGACCACCGAGATCTACA CTATGGTAATTGTGTGCCAGCM GCCGCGGTAA) and barcoded reverse primer 806R (CAAGCAGAAGACGGCAT ACGAGATnnnnnnnnnnnnAGTCAGTCAGCC GGACTACHVGGGTWTCTAAT)[21]. Each polymerase chain reaction (PCR) was prepared at ACAD, using ultraclean reagents and following strict ancient DNA protocols[22]. Each PCR tube contained 17.25 µL DNA-free water, 2.5 µL 10X ThermoPol Buffer (New England Biolabs), 0.25 uL HiFi Taq polymerase (New England Biolabs), 1.0 µL MgCL2, 1.0 µL of each primer, and 2.0 µL of genomic DNA, and each reaction was repeated in triplicate. Sealed reactions were then transported to a modern DNA laboratory at the University of Adelaide, and 16S rRNA targets were amplified under the following conditions: 95°C for 5 minutes; 37 cycles of 95°C for 0.5 min, 55°C for 0.5 min, 75°C for 1 min; and 75°C for 10 minutes. PCR products were cleaned (Ampure, New England Biolabs) to remove primers and enzymes, and DNA concentration was assessed (TapeStation, Agilent). Samples were pooled at equal nanomolar concentrations, and amplicons were sequenced on an Illumina MiSeq 2x150 bp kit using 12 bp custom indexes.

281 To determine if increasing the number of PCR cycles would increase DNA and
282 OTU yield from ancient samples, a Spy Neandertal (ACAD 14017), European
283 Mesolithic hunter-gatherer (12014 and 12017), chimpanzee (12873), and an
284 extraction blank control (14022) sample were amplified for either 38 or 43 cycles
285 under the same conditions. Increasing the number of PCR cycles significantly skewed
286 the bacterial communities observed within these samples (Figure S2). When
287 amplicons were produced with 43 cycles of PCR, the Spy I Neandertal sample were
288 dominated by Proteobacteria, and bacteria within other phyla became completely
289 undetectable, such as Bacteriodetes (Figure S2). Although more DNA was amplified
290 when the number of PCR cycles was increased, this analysis suggests that PCR bias
291 due to increased amplification can significantly affect metagenomic analysis during
292 ancient DNA studies, as expected[23,24]. Although increasing the number of PCR cycles
293 may be required to amplify DNA from ancient samples of decreased quality and
294 preservation, the minimum number of required cycles should always be utilized to
295 decrease PCR bias.
296 *16S rRNA Amplicon Bioinformatic Analysis*
297 Two independent sequencing runs generated a total of 14,644,648 DNA
298 sequencing reads. Sequences were demultiplexed to sample-specific fastq files from
299 input BCL files using the Illumina CASAVA pipeline (version 1.8.2). Overlapping
300 forward and reverse reads were joined (based on a maximum of 5% nucleotide
301 difference over a minimum 5 bp overlap) using fastq-join[25], and singletons remaining
302 after joining were discarded. Sequences were then trimmed using CutAdapt[26], and
303 subsequent files were converted to QIIME-formatted fasta files using a publically
304 available script from G. Watts (available here:
305 http://www.u.arizona.edu/~gwatts/azcc/QIIMEfastaFormatter.pl). An average of

246,427 sequences/sample were then uploaded into QIIME (MacQIIME v1.5.0), a bioinformatics pipeline-based software for the analysis of metagenomic data[27]. OTUs were determined by clustering at 97% similarity in UClust[28], and representative sequences (first or cluster seed) were selected for each cluster. By default, clusters with less than five sequences were eliminated from the analysis. Lastly, 16S sequences were given taxonomic assignments using the Greengenes database (v13) with the default setting (0.8)[29,30]. After singletons were removed, over 29,000 OTUs were observed using this approach prior to filtering, indicating that DNA damage and contamination could be inflating the number of observed OTUs, as expected[31].

Many of the ancient individuals analysed in this study, including the Spy and El Sidron Neandertals, historic primates, and ancient humans, have been stored in museum collections for long periods of time. In addition to microorganisms introduced through the soil, it is also possible that microbial contaminants are introduced during museum handling and storage or during site excavation. For example, Spy cave was excavated in 1886, and since that time, nearly all of the biological samples and archaeological remains from that site have been stored in museums, handled by scientists, and exposed to numerous storage boxes and facilities. Sequencing soil from Spy cave to analyse environmental contaminants is also is also problematic, because there are no available soil samples from the layers where the bones were collected due to the considerable length of time that has passed since the excavations. However, mammalian bones from the Spy site have been tested for DNA leaching from other bones in the site, and no evidence was found for DNA leaching within the cave[32]. Environmental exposure, sampling handling, and storage are significant issues when conducting ancient metagenomic research, and in this study, we have applied all currently available precautions to eliminate and examine

this contamination within these ancient metagenomes. However, further assessment criteria, improved laboratory methods, and new bioinformatic tools should be developed to aid in this process.

Therefore, we considered and accounted for two types of contamination in our downstream analysis for both the 16S and shotgun data sets: environmental/museum contamination and laboratory/reagent contamination. To account for environmental and museum contamination, we decontaminate the outer surface of the sample by both UV treating the sample and decontaminating this in bleach. To assess the contributions of laboratory contamination, we filtered the datasets for any taxa (eukaryotic or microbial) that are observed in soils that are analogous to that at archaeological sites. Previous publications have shown that environmental contamination, if present, leaves a unique signal microbial within the calculus[1,31]. To some extent, this signal can be identified and either analysed or removed from the dataset.

An initial analysis of 16S OTUs revealed taxa associated with laboratory contaminants and the environment were present in all calculus samples, including modern calculus samples. As is common in ancient DNA analysis protocols[33] and as has been recently recommended by Salter *et al*.[34], a conservative filtering regime was applied to identify and remove all of these contaminants from each sample. Extraction blank controls (EBCs) were included during the DNA extraction steps. EBCs were generated by including blank tubes during the demineralization step and were subjected to all of the reagents and processes that each ancient DNA sample undergoes. Therefore, sequences in EBCs are indicative of contaminant DNA present in the laboratory environment, reagents, and plasticware. Two EBCs were included for each DNA extraction that occurred for this project. Sequences from each EBC

356 also underwent the same quality filtering, trimming, and OTU (taxa) selection process
357 as reads present in ancient and modern DNA samples. Taxa identified within EBCs
358 were present in higher proportions in ancient samples, likely due in part to the age and
359 preservation of DNA within those samples (Table S2). After OTUs had been
360 identified for all the samples and EBCs collectively, any OTU identified in any EBC
361 sample was removed from the ancient and modern samples. For example, if there
362 were 10 reads for OTU 1 in Sample 1, and 5 reads for OTU 1 in an EBC sample, all
363 10 reads matching to OTU 1 were removed from Sample 1, as well as any other
364 sample containing OTU 1. Because this step was done post-OTU identification, whole
365 OTUs were removed, rather than the number of reads. If a specific *Staphylococcus*
366 *aureus* OTU was identified in an EBC, then the specific OTU that matched that *S.*
367 *aureus* OTU was filtered out from the ancient and modern calculus samples. This
368 does not mean that all OTUs identified as *S. aureus* were removed.
369 Each of the extraction blanks contained different OTUs, likely stemming from
370 the different extraction dates over the course of this project. Nevertheless, the six
371 most abundant OTUs were classified as Comamondaceae, Pseudomonadaceae,
372 *Acinetobacter, Comamonas*, *Tepidimonas*, and *Burkholderia* (Figure S3), and were
373 similar to published findings[34]. The removal of the OTUs from EBCs within each
374 sample removed an average of 87.2% of the total reads from each ancient 16S rRNA
375 calculus sample data set, suggesting that ancient samples contain large levels of
376 laboratory contamination (Table S2). Although the quantity of reads removed from
377 the calculus samples was high, this was not unexpected[34]. This is typical within
378 ancient DNA studies, as background contamination from laboratory reagents can
379 swamp out the ancient sequences available within an ancient DNA extract.
380 Laboratory contaminants can similarlly plague modern amplicon based metagenomic

studies, as an increase in endogenous signal was observed when OTUs from EBCs were removed from an analysis[35]. In this study, removing EBC OTUs is an essential step to remove the noise and increase the endogenous signal for downstream phylogenetic and comparative analysis, especially within ancient samples that can be easily overwhelmed with modern DNA contaminants.

Next, to address possible environmental or soil contamination in ancient samples that have been buried or recovered from the ground, OTUs identified in temperate coniferous forest soil were additionally removed. Six fna files corresponding to coniferous forest soil samples were downloaded from MG-RAST (mgrast_ID: mgp72) and analysed in parallel with calculus samples[36]. These soil metagenomes were selected for experimental continuity (*i.e.* amplification of identify 16S V4 region) and similarity to the archaeological sites examined in this study (*i.e.* temperate forest), as no experimentally similar cave soil data sets were available when this analysis was completed. OTUs from the soil data set were extracted and subsequently filtered from the already EBC-filtered data, as earlier described. On average, 1% of OTUs corresponding to soil were removed from the EBC-filtered data (Table S2), suggesting the ancient samples are not highly contaminated with environmental soil DNA or, more likely, that strict sample decontamination procedures prior to extraction minimize soil DNA contamination.

Finally, common laboratory contaminants were also removed. Several bacterial genera have been repeatedly detected within bacterial DNA extraction kits, enzymes, and reagents[34]. Therefore, these genera were additionally filtered from our analysis: *Acidovorax, Acinetobacter, Aeromicrobium, Afipia, Bacillus, Bradyrhizobium, Brevibacillus, Brevibacterium, Burkholderia, Caulobacter, Delftia, Devosia, Herbaspirillum, Lamia, Mesorhizobium, Methylobacterium, Ochrobactrum,*

406 *Olivibacter, Paenibacillus, Pedobacter, Propionibacterium, Pseudomonas,*
407 *Pseudoxanthamonas, Ralstonia, Rhizobium, Rhodococcus, Sphingomonas,*
408 *Stenotrophomonas,* and *Sulfuritalea*. While most of these OTUs had been removed in
409 the previous EBC OTU filtering step, on average an additional 5.3% of the EBC- and
410 soil-filtered reads were removed. The number of reads removed from each sample
411 was higher in the ancient samples, as expected (Table S2). After this filtering, each
412 sample contained an average of 19.4 thousand reads (Figure S4).
413 Upon removal of environmental and laboratory contaminants, there were
414 several samples with lower numbers of reads (<500 reads) that also contained only a
415 few bacterial phyla (<3 phyla). The low bacterial diversity likely signifies that the
416 remaining reads do not represent an intact oral bacterial community. Therefore, four
417 samples were excluded from downstream analysis, which resulted in the removal of
418 three South African individuals (AfrPP3, AfrSF1, AfrSF5) and the baboon.
419 Nevertheless, over 775 different OTUs were detected using this method, and 30
420 different OTUs were present within all of the cultures examined. In addition, filtered
421 modern calculus samples were comparable to oral bacterial communities identified by
422 Adler *et al*.[37] and the Human Oral Microbiota Database (HOMD)[38], while remaining
423 distinct from environmental controls, including soil, sediment, water, and an
424 archaeological tooth sample (Figure S5).
425
426 ***Shotgun Metagenomic Library Construction and Analysis***
427 *Shotgun Metagenome Library Construction*
428 Shotgun metagenomic libraries for new samples collected in this study were
429 constructed as previously described for ancient DNA research[20,39], with the following
430 modifications. Briefly, 20 μL of DNA was polished in 20 μL reaction with T4

polynucleotide kinase (New England Biolabs) and T4 DNA polymerase (New England Biolabs) for 15 minutes at 25°C. Reactions were cleaned using a MinElute Reaction Cleanup kit (Qiagen). Truncated Illumina adapter sequences with 5 bp unique barcodes[39] were ligated onto double stranded DNA molecules using T4 DNA ligase (Fermentas) for 60 min at 22°C. After a clean-up using a Qiagen MiniElute Reaction Clean-up kit to removed excess ligase, adapter sequences were filled using a Bst DNA polymerase (New England Biolabs) for 30 minutes at 37°C, followed by denaturation of the polymerase at 80ºC for 10 minutes. The resulting reaction was then used as a template in five independent PCR reactions (12 µL DNA-free dH_20, 2.5 µL 10x buffer, 2.5 µL 25 mM $MgCl_2$, 0.25 µL 25 mM dNTPs, 1.25 µL of IS7 and IS8 primer sequences[39], 0.25 µL HiFi DNA polymerase, and 5 µL of the Bst reaction mixture) under the following conditions: 12 min at 94°C; 13 cycles of 30 sec at 94°C, 30 sec at 60°C, 45 sec at 72°C; and 10 min at 72°C. PCR reactions were pooled and cleaned using Ampure PCR purification (Agencourt). Libraries were then re-amplified with the same conditions using GAII indexing primers[39] to include a single P7 (3') index sequence unique to each sample, re-pooled, and cleaned again using Ampure. Libraries were quantified using a TapeStation and quantitative PCR (KAPA Illumina quantification kit) and pooled at equimolar concentrations prior to paired-end sequencing on an Illumina HiSeq (Neandertals; chimpanzee; modern human) and NextSeq (all other remaining samples).

Shotgun Metagenomic Bioinformatic Analysis

Raw fastq files obtained from either sequencing machine were demultiplexed using Sabre 1.0 (available here: https://github.com/najoshi/sabre) according to sample specific index sequences, and reads were then merged using bbmerge (available here: http://sourceforge.net/projects/bbmap/). Adapter removal was then used to identify

456 reads that matched both the 5' and 3'barcode sequences (5 bp) and trim both the
457 barcode and adapter sequences from the reads. Taxonomic identifications of collapsed
458 (merged) reads were identified using MALTX, as non-merged reads were greater than
459 300 bp in length and likely represent more contamination than endogenous signal.
460 Briefly, MALTX (v0.0.12) was developed by the Huson lab at the University of
461 Tuebingen and is an in-house Java program that compares DNA reads against a
462 protein reference database, such as NCBI-nr[40]. It employs a seed-and-extend strategy
463 using spaced seeds and a reduced alphabet in the seed step, similar to DIAMOND[41].
464 MALTX employs the same spaced seeds, reduced alphabet, and alignment strategies
465 as DIAMOND but uses a hash-table rather than double-indexing to find seed matches
466 between queries and references. The output of both programs is similar, but not
467 identical, due to differences in the heuristics used. MALTX output files were
468 converted into rma files and uploaded into MEGAN5[42]. Species identified in EBC and
469 environmental (water and soil) controls were removed from calculus samples in
470 MEGAN5; sequences identified at higher-level classifications in the control samples
471 were not removed from the data, due to unknown provenance. Briefly, species
472 identified in EBCs or environmental controls (Figure S16) were selected
473 (Select>Leaves) from all calculus samples. The inverse of these species (*i.e.* all taxa
474 not selected) were then exported into a new file, conserving all species not identified
475 in controls and any higher order taxa. This data file was then utilized for downstream
476 analysis.
477 In MEGAN5, alpha diversity was conducted on the filtered data set by
478 calculating both Simpson's and Shannon's inverse indexes from all taxa, while beta
479 diversity was examined by calculating the Euclidean and Bray-Curtis distances of the
480 genera identified in each sample. UPGMA clustering of distances was performed in

481 SplitsTree[43], and tree construction was visualized and edited in FigTree v1.4.1
482 (available at: http://tree.bio.ed.ac.uk/software/figtree/). GraftM was employed to
483 identify sequences corresponding to the 16S ribosomal RNA encoding genes
484 (available at: https://github.com/geronimp/graftM). These sequences were selected by
485 a hidden Markov chain model in hmmer against the RDP database to ensure
486 eukaryotes were distinguished, and identified by their placement into a phylogenetic
487 tree using pplacer. Metagenomic and 16S rRNA compositions were compared to
488 GraftM analysis by normalizing the taxonomic classifications across NCBI, RDP, and
489 Greengenes identifications to the NCBI convention during downstream analysis in
490 MEGAN5. Statistical analysis of shotgun data was completed using LefSe[44].
491 *Genomic and Phylogenetic Analysis*
492 Patterns of DNA damage were plotted from shotgun sequencing for several
493 species. First, species of interest, including modern oral, respiratory, and gut
494 pathogens, were identified from the MALTX output. Collapsed reads from each
495 species were then mapped to the identified reference genome with BWA v0.6.2 [45],
496 using the parameter space recommended for ancient DNA (no seed, one gap opening,
497 relaxed edit distance)[46]. Duplicate reads were removed using
498 FilterUniqueSAMCons.py[47]. For example, MALTX identified the highest number of
499 archaeal sequences as matching to *Methanobrevibacter*, so both the gut and oral
500 *Methanobrevibacter* genomes were used independently as a reference sequence to
501 map any and all reads from fastq files. Reads aligned to either genome were then
502 analysed in MapDamge 2.0.2[48]. Once the correct reference was identified (*i.e.*
503 whichever reference genome matched the most sequences), the identification of genes
504 present in the modern genome, but putatively absent in the ancient genome (*i.e.* no

505 reads mapping to the reference loci), were completed by comparison in Geneious[49].
506 Plots of G/C content and mapped genomic regions were constructed using Circos[50].
507 Next, phylogenetic relationships between ancient bacterial species and modern
508 relatives were assessed. All available published genomes and whole genome
509 sequences of the identified species and closely related taxa were obtained from NCBI,
510 and aligned using default parameters in progressiveMauve[51]. Outputs files were
511 converted into a fasta format using bx-python scripts (available at:
512 https://github.com/bxlab/bx-python). Dissimilar and poorly aligned sections from
513 these genomes were removed using GBlocks[52]. Phylogenetic trees of the aligned
514 sequences were estimated using the maximum likelihood (ML) procedure in RAxML
515 v 8.1.21[53] using the GTRGAMMA substitution model. The best ML tree was
516 retrained from 20 ML computations, and support values were obtained from 100
517 bootstrap replicates using the rapid bootstrapping algorithm[54]. Divergence times
518 between closely related strains (*i.e. M. oralis* and *M. smithii*) were estimated using the
519 Bayesian Markov chain Monte Carlo procedure available in the BEAST package[55].
520 We used both strict and relaxed (uncorrelated lognormal) molecular clock models,
521 and we assessed sufficient sampling by verifying that the effective samples size of all
522 parameters was at least 200. Finally, we calculated the ratio of non-synonymous to
523 synonymous nucleotide substitutions per site (d_N/d_S) for the ancient and modern *M.*
524 *oralis* strains. We selected the genes that corresponded to protein coding regions and
525 transcribed them using the bacterial, archaeal, and plant plastid code as implemented
526 in BioPython[56]. If stop codons were identified in the ancient strain, which might occur
527 due to reading frame shifts or particular features of the archaeal genome, the coding
528 sequence was removed. Of the 1990 gene identified, 375 genes met these criteria. We
529 then used a custom script (available at https://github.com/sebastianduchene/adna_dat)

530 to calculate the number of non-synonymous and synonymous substitutions, as our
531 analysis was restricted two sequences (the only available reference sequence and the
532 ancient genome; raw count), which precludes the use of more sophisticated codon
533 substitution models. A cut-off value of 0.1 was utilized to identify genes under
534 purifying selection, while a cut-off of 1 was utilized for genes under positive
535 selection.

536 *Dietary Analysis*

537 The shotgun data set was also examined for eukaryotic sequences that could
538 be indicative of dietary food sources. While eukaryotic sequences represented a small
539 fraction of the total reads, sequences pertaining to food sources could still be
540 identified. However, several initial results seemed spurious (*i.e. Drosophila* or
541 *Xenopus* in the modern human). To remove spurious results, we removed reference
542 genomes that contained known levels of human DNA contamination[57,58]. This filtered
543 many of the spurious results with the exception of ticks in the modern human (Table
544 1). The tick genome was not included in published analyses that examined human
545 DNA contamination within genomes, so the validity of this finding is insecure. It is
546 highly unlikely that this modern individual was eating ticks, and more likely
547 represents unidentified contamination in the tick genome that originates from either
548 human or microbial DNA. Similarly, the Spy I Neandertal also contained hits to many
549 spurious results, which also likely arise from modern contamination. Regardless,
550 dietary analysis from dental calculus is a burgeoning new field, and accurate reference
551 genomes that reflect ancient taxa are very limited. As damage patterns could not be
552 investigated due to the limited number of hits, we must note the limitations of the
553 current approach. In this study, we have assumed that hits to known related modern
554 species equate with related ancient taxa. However, this is a large assumption, and

555 further investigation using enrichment hybridization techniques or deeper sequencing
556 will help verify these results. There is also always a possibility that these results
557 represent contamination, resulting from DNA leaching in groundwater, sample
558 mixing during archaeological excavation, or DNA contamination from the adhesive
559 used to assemble ancient skeletons. Certainly, further research and development of
560 dietary calculus tools are needed to fully verify the limited dietary findings reported in
561 this study.

562 ***Results***

563 **Amplicon Sequencing Analysis**

564 *16S rRNA SourceTracker Analysis*

565 To explore bacterial contaminants and the impact of quality filtering on the16S
566 rRNA data set, the raw (pre-filtering) and filtered sample OTU tables were analysed
567 using the take-one-out method in SourceTracker 0.9.6[59]. Comparison samples (human
568 skin and gut, soil, outdoor air, and indoor air microbiota) available with the Source
569 Tracker package were used to identify the source of contamination introduced into
570 ancient samples. To ensure EBC and negative-template PCR control contamination
571 was effectively removed during the filtering process, EBC and PCR negative samples
572 were also used as comparison samples. OTUs present in less than one percent of
573 samples were also removed to increase computational speeds[60], and sample groups
574 (cultures) that contained only one sample were excluded in the take-one-out analysis.
575 The results were averaged across all samples per group, and the proportions were
576 attributed to non-oral human microbiota, soil, and indoor/outdoor air. OTUs falling
577 within the SourceTracker default "Unknown" category were collapsed into a single
578 value ("Other").

579 As expected, the raw data (Figure S6A) showed a mixture between cultures,
580 compared to the filtered data (Figure S6B). This indicates shared sequences exist
581 between all samples in the raw data and likely represent laboratory contamination.
582 Indeed, these sequences were removed during the filtering process. Hence, filtering
583 enhances signal within groups and highlights the importance of tracking and
584 removing contaminating sequences from sequencing data. It is interesting to note that
585 more ancient/poorly preserved samples (Neandertal and African samples) contained
586 increased proportions of contaminant reads attributed to the EBC and PCR negative

587 samples, likely due to preservation bias. Higher contamination levels would be
588 expected in older samples, because the endogenous DNA proportion typically
589 decreases with samples age[61]. As expected, the filtered data shows no overlap of
590 sequences between the EBC and PCR negative samples with the calculus samples.
591 Additional non-oral bacterial communities were collapsed into a single category
592 ("Other"), as no calculus sample (raw or filtered) had derived sequences from these
593 groups. This demonstrates that the non-oral bacteria identified in ancient samples are
594 unique and are not large components of the human mouth, skin, or the environmental
595 contamination examined here.
596 *16S rRNA Dissimilarity Analysis*
597 To identify unique OTUs and determine how they may contribute to
598 community-based phylogenetic analysis, dissimilarity indexes were first calculated
599 and compared for OTUs falling within the 99th percentile. Filtered OTU tables from
600 the calculus samples and EBCs were individually converted into a matrix using an in-
601 house python script and R[62]. Raw read within samples were counted, and the
602 maximum relative abundance was determined. The data sets were then filtered for
603 taxa that contained a proportion of >1% abundance, and the Jaccard dissimilarity
604 distance was calculated using the vegan package within R. Jaccard distance is based
605 on presence-absence similarity, in contrast to the abundance related measurements,
606 such as Bray-Curtis and Chao indices[63]. The abundance for each taxa >1% was
607 organized according to Jaccard similarity and placed in a heat map (Figure S7; Table
608 S3). Taxa present at greater than 5% abundance were also identified and flagged for
609 display purposes as highly abundant OTUs. All R scripts to create heat maps and
610 Jaccard dendograms are available upon request.

611 Utilizing these heat maps and Jaccard matrixes, four clusters were identified,
612 which have been simplified as: No-agriculture (hunter-gatherers, foragers, and some
613 pastoralists); Agriculture (early farming cultures); 19th Century (Industrial
614 Revolution); or Modern (Figure S7). Interestingly, a hunter-gatherer sample
615 (EuroHG-2) fell intermediate between No-agriculture and Agriculture groups (Figure
616 S7; Table S4) due to a single OTU (*Nitrospirales* sp.), which may reflect contact with
617 contemporaneous agriculturalists or the transition to an agriculturalists based diet[64,65].
618 This may also demonstrate the lack of reproducible signal when analysing ancient
619 amplicon data sets. This analysis highlighted several notable phyla-level patterns
620 when the OTUs present at >1% were grouped into their respective phyla (Figure S12;
621 Table 4). Firmicutes and Proteobacteria were typically identified in ancient non-
622 agriculturalists, while an increase in Bacteroidetes was observed in more modern
623 samples (IndRev and modern). Fusobacteria were only identified as highly abundant
624 in modern individuals. While potentially biological, these large-scale differences also
625 appear to correlate with the age of sample and again highlight potential issues with
626 utilizing >200 bp amplicons to analyse the genetic diversity within ancient samples.
627
628 **Shotgun DNA sequencing analysis**
629 *Benchmarking MALTX Analysis*
630 Species identification from diverse ancient metagenomic sequences is a
631 difficult task, especially with fragmented and damaged DNA reads. While many
632 ancient DNA studies have used BLAST-based metrics for species identification[66],
633 large data sets require analytical software with rapid, accurate analysis methods. To
634 examine how well MALTX performs relative to other rapid metagenomic analysis
635 tools (*i.e.* benchmarking), we created several *in silico* data sets from 49 bacterial

636 genomes of representative taxa previously detected in ancient dental calculus
637 genomes[1] (Table S4). To assess how MALTX would respond to short read lengths
638 versus damage introduced into ancient shotgun libraries, we simulated modern or
639 ancient (damaged) metagenomes using an average of 1 million reads per reference
640 genome at three average size length distributions (40 bp, 80 bp, and 120 bp),
641 generating six total simulated metagenome data sets. Modern simulated metagenomes
642 were constructed using ArtificialFastqGenerator[67] that simulated standard sequencing
643 errors. These simulated data sets were then subjected to damage (*i.e.* generated
644 damage within the data set) using an in-house program (SIMWRECK;
645 https://github.com/mtrw/simwreck) with the following parameters: p=3.35; d=0.3;
646 D=0.65. These simulated metagenomes were then analysed using default parameters
647 in MG-RAST[68] (best hit classification), DIAMOND[41], MetaPhylAn[69], and MALTX.
648 Simulated metagenomes could not be analysed in some situations due to size
649 fragment cut-offs (*i.e.* MALT and DIAMOND (40 bp) and MG-RAST (70 bp)).
650 Species identifications and community structures were influenced by analysis
651 method, as expected[70]. In UPGMA clustering of Bray-Curtis distances, modern and
652 simulated damaged metagenomes analysed in DIAMOND and MALTX clustered
653 more similarly for all lengths, likely reflecting similarities in the heuristic algorithms
654 employed in each method (Figure S8). MetaPhlAn results clustered to the exclusion
655 of DIAMOND and MALTX outputs, and MG-RAST outputs clustered to the
656 exclusion of all other data sets. MG-RAST analysis of ancient fragments was highly
657 biased, and resulted in data that was indistinguishable across ancient and modern
658 samples. Worryingly, the EBC samples also looked similar to modern calculus
659 samples when analysed by MG-RAST (Figure S9). MG-RAST analysis is inherently
660 limited by a 70 bp threshold for species identification and is therefore problematic to

661 analyse ancient DNA data sets, as the average sequence size in the ancient data sets is
662 less than 70 bp (Table S6). With the exception of the MG-RAST analysis that is
663 biased towards larger fragments (>70 bp), taxonomic profiles constructed from each
664 data set demonstrated similar trends (Figure S9). Interestingly, DNA damage had a
665 larger impact on taxonomic identifications than DNA fragment length (Figure S9),
666 although fragments <80 bp could not be included. DIAMOND and MALT were more
667 accurate at maintaining taxonomic identifications even when damage was included in
668 the analysis, in contrast to profiles generated from MetaPhylAn, which were highly
669 impacted by damage and clustered separately rather than together during UPGMA
670 clustering (data not shown). Overall, DIAMOND and MALTX better reflected
671 unbiased bacterial community structure that was more similar to the input species and
672 was more robust to biases introduced through different DNA fragment lengths and
673 damage.

674 In addition, raw data for all of the available modern oral metagenome data sets
675 (amplicon and shotgun) on the MG-RAST repository were downloaded and analysed
676 by MALTX to identify similarities and differences between this analysis and
677 published findings. Amplicon and shotgun genera were UPGMA-clustered according
678 to Bray-Curtis distances (Figure S10). Modern calculus deeply sequenced in this
679 study (described below) possessed bacterial phyla similar to plaque shotgun
680 metagenome samples[71]. All plaque and calculus samples clustered to the exclusion of
681 a tracheal metagenome[72], as expected. Salivary amplicon data analysed by MALTX
682 were distinct from shotgun data, as expected; however, salivary data analysed by
683 MALTX was similar to the common salivary microbiota and was stable through time,
684 as expected[73]. Overall, MALTX was able to reconstruct published results,
685 demonstrating its accuracy for diverse and ancient metagenomic samples.

686 *Impacts of Filtering and Analysis Method on Shotgun Data*

687 DIAMOND and MALTX (0.0.12) analyses yielded similar results (Figure S8),
688 but filtered DIAMOND data sets tended to have increased Proteobacteria proportions
689 even in modern specimens (Figure S18 and S19). Because the majority of taxa within
690 the EBCs were Proteobacteria, it is possible that laboratory and environmental
691 contamination were not accurately removed using species-level DIAMOND
692 identifications. Only sequences identified to the species level can be filtered from
693 shotgun data, so the inability to identify specific species in either analysis method
694 would allow more contaminant reads to infiltrate the data. Therefore, MALTX
695 analysis was selected as the best approach and was utilized for all downstream
696 analysis.

697 In addition to removing species identified in EBC samples, we also filtered
698 environmental species to assess environmental contamination and provide a primitive
699 way of estimating the 'endogenous' DNA present in a sample. We removed species
700 found in soil and water metagenomes from previously sequenced data sets available
701 in MG-RAST (MG-RAST IDs: 4536380.3, 4536373.3, 4516952.3, 4511193.3,
702 4477876.3) (Figure S15, S16, and S17; Table S7). As stated in the main text, the Spy
703 Neandertal samples were drastically impacted by the lab and environmental filtering
704 (80-94% reads removed). This was in stark contrast to the El Sidrón samples (12-
705 29%), indicating that less contaminant DNA is present in the Spanish Neandertal
706 material. In anatomically modern human (AMH) samples, the impacts of filtering
707 were independent of sample age, as ~8,000 year-old AMH European hunter-gatherers
708 had no sequences removed, while contemporaneous European LBK samples had up to
709 70% of the sequences removed (Table S7). The extraction facility also appeared to
710 play a role in the level of background contamination present in the samples. For

example, two Medieval calculus samples processed by Warinner *et al*., contained >99% background sequences, whereas the modern sample extracted in a modern DNA facility at the University of Adelaide, Australia, had 39% of the sequences removed from the data. The youngest samples processed in the ancient facility (Industrial Revolution samples) both contained <2% background contamination. Overall, this filtering process estimated that up to 24% of the total ancient sequencing reads identified in this study are non-endogenous, in stark contrast to the 82.7% identified in amplicon data sets.

Diversity Analysis of Shotgun Data

To analyse species diversity through time, Shannon and Simpson-Weaver indexes were calculated from filtered (non-rarefied) shotgun data sets in MEGAN5 (Figure S20). An earlier examination of Gram-positive microorganisms from ancient dental calculus identified a statistical decrease in species diversity over time[1], and decreased diversity has been reported in several studies examining the microbiota of greater apes compared to modern humans[74]. Statistical differences in diversity through time were not observed between samples of this study, although differences were observed in samples from this study compared to others from Warinner *et al.*[31] (Figure S20). Due to the large differences in sequencing depth between these two studies, species diversity was likely impacted by the number of raw sequences obtained from each sample. This observation, paired with the large levels of environmental and background DNA contamination, highlights the difficulties in using alpha-diversity metrics to accurately assess diversity in ancient samples.

Differences between samples (beta-diversity) were examined by calculating Bray Curtis distances between all samples within the rarefied (*i.e.* an equal number of sequences or all taxa identified) and complete data sets. Pairwise Bray Curtis

736 distances in MEGAN5 were calculated from the genus level subtree of Bacteria and
737 Archaea. UPGMA trees were built using these distances in SplitsTree and visualized
738 in FigTree. Within the complete data set, the analysis identified four distinct clusters
739 (Figure 2B and Figure 21A). Each of the four clusters is described in the main text
740 and can be easily differentiated by the large-scale changes in Gram-positive/negative
741 proportions (Figure S21B). The rarefied data revealed similar placement of European
742 samples, but the placement of the African individuals varied between the two
743 approaches (Figure S21A). The rarefied data was limited to only 340 sequences per
744 sample, as some of the African individuals contained few identifiable sequences.
745 Rarefying the data to this level is likely inappropriate and may bias the clustering
746 (Figure S21A). For example, the complete data set revealed an increase in support for
747 the identified nodes (*i.e.* nodes values in Figure 2B compared to Figure S21A), and
748 samples did not cluster based on sequencing depth when the complete data set was
749 analysed. With the complete data set, the placement of non-African samples
750 remained consistent, suggesting that depth does not contribute to their placement on
751 the tree. Therefore, we chose to compare the differences between each of the four
752 main clusters identified within the complete data set in the main text, utilizing all of
753 the limited taxa identified for the African samples for their placement on the tree.
754 There are several interesting observations within these four groups.
755 Surprisingly, one LBK individual (Tubingen collection 1979, grave 111) clustered
756 with Mesolithic and Para-Neolithic European hunter-gatherers (6645 +/- 30 BP (Late
757 Mesolithic) to 4690 +/- 40 BP (Para-Neolithic, Zedmar culture)) and the Spy II
758 individual. Misidentification or cultural admixture does not likely explain this
759 observation, as this LBK individual possesses an H mitochondrial DNA haplotype[75],
760 as expected in early European farmers. While speculative, Neandertal and AMH

761 interaction, climatic alterations, or unknown biocultural changes could have all
762 contributed to this observation. Ancient calculus samples from another study by
763 Warinner *et al.*[31] also cluster with the agriculturalist group, as expected, reflecting the
764 robustness of this approach. Despite these differences, the separation of ancestral
765 hunter-gatherers and modern individuals from agriculturalists is clear.
766 Despite these large-scale differences, several species were also shared across
767 multiple samples. For example, 18 different species were conserved across
768 chimpanzees, Neandertals, and modern humans (Table S8), and each contained
769 damage patterns consistent with sample age (Figure S22), indicating several oral taxa
770 that have been conserved over evolutionary time. The average DNA damage in these
771 conserved bacterial species was less in the El Sidrón Neandertals (31% C-to-T to 36%
772 G-to-A) than damage observed in the published mitochondrial DNA (mtDNA)
773 sequences from a related individual (El Sidrón 1253 mtDNA; 50% C-to-T)[76] (Figure
774 S22A), consistent with previous findings[77]. Surprisingly, apparent damage in Spy
775 bacterial species was markedly less (0.04-13% C-to-T and 11-12% G-to-A) but was
776 still demonstrably higher than that from modern specimen (3% for both C-to-T and G-
777 to-A). As mapped read lengths in Spy Neandertals were not indicative of modern
778 cross contamination (54 bp) (Figure S22B), inefficient mapping in these species could
779 contribute this phenomenon; 89% fewer reads mapped in Spy Neandertals compared
780 to El Sidrón Neandertals due to low library complexity. In comparison, the average
781 read length for bacterial species in El Sidrón Neandertals appeared ancient (50 bp)
782 and was similar to the length of published mtDNA sequences (52 bp). Damage of
783 mitochondrial DNA within ancient calculus could not be assessed, as minimal
784 Neandertal mitochondrial DNA sequences were identified (El Sidrón 1:5; El Sidrón
785 2:0; Spy 1:56; Spy 2:0; <0.00004% of the total reads), although the identified

786 sequences were longer than expected (~82.3 bp) (Table S13). In contrast, 9.4-fold
787 coverage (99.6%) of a human mitochondrial genome could be reconstructed from
788 modern calculus (1.5% C-to-T damage) (Table S13). Despite studies that have applied
789 hybridization enrichment techniques to obtain mtDNA sequences from historic dental
790 calculus[78], these results suggest that human DNA is not a substantial component of
791 ancient dental calculus, suggesting that human DNA may be excluded or degrade
792 more rapidly because it is not a structural component of the bacterial biofilm.
793 *Specific Pathogen Identification and Analysis*
794 Once species-level identifications were obtained using MALTX, specific oral
795 and respiratory pathogens of interest were examined more closely. To ensure
796 authenticity of pathogen identifications, at least four of the following five criteria
797 were required:
798 1.) Several genes from the pathogen must be identified in sequenced data (i.e. >20
799 genes identified in MALTX results).
800 2.) Genome-wide reference-mapping and damage analysis must reveal DNA
801 damage patterns indicative of ancient DNA fragments, *i.e.* >500 reads
802 mapping to a reference genome with C-to-T rates expected for thermal age.
803 3.) The pathogen cannot be detected in extraction blank control samples or
804 environmental samples (*i.e.* soil) that was analysed using similar methods.
805 4.) The pathogen of interest cannot be closely related to common commensal oral
806 isolates, *i.e.* its lineage must pertain to a unique branch on a phylogenetic tree.
807 5.) Skeletal or paleopathological evidence should exist within the population to
808 support the existence of the disease caused by the pathogen.
809 Each of the pathogens identified within this study (Tables S9 and S10) were
810 scrutinized using these criteria. These criteria provide a conservative guideline for

811 pathogen identification but cannot differentiate between progenitor strains and current
812 pathogenic strains in ancient samples. In the case of several pathogens identified in
813 ancient samples, their identifications may be attributed to progenitor or closely related
814 ancestral strains of modern day pathogens. This method is also unable to accurately
815 predict if individuals suffered from the disease or simply carried the pathogen
816 asymptomatically.

817 Several modern pathogens were identified in ancient samples, but were also
818 identified in environmental control samples (Table S9 and S10). The caries-associated
819 pathogen *Streptococcus mutans* was identified in all Neandertals (0.08% to 0.18%), as
820 well as fresh ground water. All three members of the 'red complex' pathogens
821 associated with periodontal disease were identified in Neandertals (*Porphyromonas*
822 *gingivalis*: 0-0.52%; *Tannerella forsythia* 0.05-2.4%; and *Treponema denticola* 0-
823 1.87%) and fresh ground water (Table S9). It is unlikely that these obligate human
824 pathogens from the oral cavity can survive in the environment, so we examined if
825 these oral species were present in common laboratory reagents. *T. denticola, P.*
826 *gingivalis,* and *T. forsythia* were all identified in a common DNA extraction kit (Table
827 S9). While environmental samples were extracted using kits, the ancient samples in
828 this study were extracted using homemade QG silica method that is free of these
829 pathogens, as demonstrated by the EBCs extracted simultaneously with ancient
830 samples (QG EBCs 1-2) (Table S9). Further, DNA from these pathogens in
831 Neandertals contained damage indicative of ancient DNA (Table S11). Together,
832 these data indicates that Neandertals shared these pathogens, or at least ancestral
833 forms or genetic content of these pathogens, with modern humans.

834 This approach also identified five nasopharyngeal and respiratory pathogens in
835 Neandertals that were absent in all controls and environmental samples: *Bordetella*

836 *parapertussis* (whooping cough), *Pasteurella multocida* (skin and respiratory
837 infections), *Streptococcus pyogenes* (strep throat), *Corynebacterium diptheriae*
838 (diphtheria), and *Neisseria gonorrhoeae* (gonorrhoea) (Table S9). In El Sidrón
839 samples, each of these pathogens contained DNA damage indicative of ancient DNA
840 (Table S12). The first four of these are important nasopharyngeal and respiratory
841 pathogens of modern humans, but *N. gonorrhoeae* does not typically infect the oral
842 cavity. Several of these pathogens (*N. gonorrhoeae*, *S. pyogenes*, and *C. diptheriae*)
843 are also closely related to common oral microorganisms, which confounds accurate
844 identifications in ancient metagenomic samples and therefore disqualifies them from
845 further identification using our methods. For example, genomic mapping and
846 phylogenetic analysis of reads identified to *N. gonorrhoeae* in El Sidrón 2 placed this
847 strain within known commensal oral *Neisseria* isolates (Figure S23). Small fragment
848 size typical of ancient DNA confounds accurate mapping of reads to species and
849 would also likely influence accurate functional profiling. Until enhanced mapping
850 algorithms specific to short ancient DNA molecules are developed for metagenomic
851 analysis, these types of analyses will remain difficult.
852 The application of authenticity criteria did suggest that *Bordetella*
853 *parapertussis*, a bacterial respiratory pathogen that causes whooping cough, and
854 *Pasteurella multocida*[79], a bacterial pathogen known to cause lethal co-infections with
855 whooping cough, in the El Sidrón Neandertals (Table S9) may be endogenous. In both
856 specimens, the *Bordetella* reads mapped most closely to a zoonotic isolate of *B.*
857 *parapertussis* strain BPP5 and included bordetellae specific genes (*i.e.* BPP5 fosmid
858 and filamentous hemagglutinin encoding genes, *fha*). Reads mapping to the BPP5
859 genome from both Neandertals were ancestral to all pathogenic classical bordetellae,
860 including zoonotic isolates of *Bordetella bronchiseptica* and human-restricted strains

861 of *Bordetella pertussis* (Figure S26). Currently, there are no known or closely related
862 human oral isolates of *Bordetella* or *Pasteurella*, and phylogenetic analysis identified
863 *B. parapertussis* reads as basal to many modern *Bordetella* pathogens. Despite
864 adhering to these criteria, the observation of *B. paraperutssis* in El Sidrón Neandertals
865 could be explained by either the presence of an ancestral form of *Bordetella*, *B.*
866 *bronchiseptica* present within soil at the site, or more simply by noise and inefficient
867 mapping in the data set. To test the efficiency of the mapping, we directly compared
868 mapping quality scores from reads that mapped to two genomes, identifying which
869 reads mapped more efficiently to a specific genome. Only 20% of reads that mapped
870 to *B. parapertussis* more efficiently mapped to *B. petrii*, and even fewer reads more
871 efficiently mapped to *B. parapertussis* than to a highly unrelated *Neisseria* pathogen,
872 confounding the accurate identification of *B. parapertussis* in Neandertals. Further
873 analysis and the development of more accurate mapping tools that handle short reads
874 will be required to accurately identify this pathogen. Nevertheless, the identification
875 of oral pathogens and *Bordetella* within Neandertals may indicate that these
876 pathogens have been evolving with hominins over expansive time periods and did not
877 originate during a recent zoonotic episode during animal domestication[80].
878 While Neandertals went extinct >30,000 years prior to the onset of agriculture,
879 their pathogens provide critical information to examine the history and origin of
880 infections in modern humans. The Neolithic Revolution, with large dietary shifts,
881 domestication of animals, sedentary lifestyles, and decreased sanitation[81], is estimated
882 to be the single most important event to impact human disease[82] and is believed to be
883 a significant driver of pathogen evolution[83]. However, the identification of a
884 whooping cough and several other conserved oral and respiratory bacterial pathogens
885 within Neandertals indicates that several infectious diseases, thought to have been

886 first transmitted after the onset of agriculture, existed in hominids well before this
887 transition. Hence, modern 'zoonotic' pathogens may not have been introduced into
888 humans during domestication and are potentially the result of genomic evolution or
889 different selection pressures that arose during the Neolithic Revolution. This also
890 suggests that these pathogens, or at least their ancestral forms, have been shared over
891 longer evolutionary time spans than previously estimated. Further studies aimed at
892 examining the genomic evolution through time in pathogens will likely uncover
893 molecular mechanisms of evolution responsible for the spread of infectious agents
894 during the Neolithic Revolution.

895 *Methanobrevibacter Analysis*

896 Oral *Methanobrevibacter* taxa are very similar to *Methanobrevibacter* species
897 found in the gut (*i.e. M. smithii*) and differ largely through the loss of several large
898 adhesion associated genes likely used to attach to the epithelial cells of the intestinal
899 tract. Therefore, we wanted to investigate the evolutionary history of *M. oralis* subsp.
900 *neandertalensis* further to (1) identify if archaeal DNA degrades at a rate similar to
901 bacterial DNA, (2) reveal if there were similarities to the archaeal gut relative, and (3)
902 determine when the oral and the gut *Methanobrevibacter* taxa diverged. First, we
903 mapped ancient sequences to *M. smithii* and to *M. oralis*. 79% fewer genes mapped to
904 *M. smithii* ATCC 35061 than mapped to *M. oralis* JMR01 genome, and the ancient
905 strain also lacked genes for adhesion in the gastrointestinal tract. This suggests that
906 the strain in the oral cavity of El Sidrón 1 is indeed *M. oralis*. Maximum likelihood
907 phylogenetic analysis with RAxML also confirmed this, as *M. oralis* subsp.
908 *neandertalensis* was more similar to *M. oralis* JMR01 than a wide-range of *M. smithii*
909 *strains*. The damage patterns were also consistent for a bacterial species of similar age
910 (Table 2). We also found that the rate of degradation in Gram-positive

911 (*Peptostreptococcus stomatis*; 54.6 bp; 36% C-to-T; 40% G-to-A), Gram-negative
912 (*Propionibacterium propionicu*m; 48.8 bp; 37% C-to-T; 43% G-to-A), and
913 eubacterial taxa (*Eubacterium sphenum*; 52.8 bp; 37% C-to-T; 41% G-to-A) was
914 similar to that observed in archaea (*Methanobrevibacter oralis*; 58.67; 33% C-to-T;
915 36% G-to-A), suggesting that cell wall structure does not play a role in bacterial DNA
916 preservation over long time periods. The damage observed in *M. oralis* was also
917 lower than previously observed in the mitochondrial DNA from a contemporaneous
918 El Sidron Neandertal (El Sidron 1253 mtDNA; 50% C-to-T; 52 bp)[76], as has been
919 previously observed with bacteria[84].
920 Using a tip-dated ancient genome, we wanted to examine both when oral
921 archaeal taxa diverged from related gut strains and when the Neandertal and human
922 *M. oralis* strains diverged. Initially, we found poor mixing of the Markov chain in
923 BEAST for our analyses of two *M. oralis*, four *M. smithii*, and one
924 *Methanobrevibacter wolinii* (an outgroup) strains under the uncorrelated lognormal
925 relaxed clock model. We attributed this to the fact that our data set consists of a small
926 number of taxa (*i.e.* this clock model is over parameterised). In particular, every
927 branch in the tree can have a different rate under this model. These rates are described
928 using two parameters: the mean rate across branches and the standard deviation of the
929 branch rates, both of which may be difficult to calculate for phylogenetic trees with
930 few taxa (*i.e.* less than 20 branches). Therefore, we interpreted the estimates from the
931 strict clock model. Previous studies involving dated tips have suggested that the mean
932 estimates from this model are generally comparable to those from relaxed clock
933 model[85]. To verify that these estimates were not mislead by high rate variation, we
934 also conducted our analyses in BEAST after selecting a subset of 192 genes with the
935 smallest departure from clocklike behaviour according to their coefficient of rate

936 variation. We assessed the impact of high rate variation among lineages in our
937 estimates by repeating our analyses in BEAST with these 192 genes with the strongest
938 clocklike behaviour, and the analysis produced largely congruent estimates of
939 evolutionary rates and divergence times. Therefore, a strict clock was utilized to
940 estimate dates provided in this study. While the date estimate for the Neandertal and
941 human *M. oralis* strains is discussed in the main text, the data for the split between *M.*
942 *oralis* and *M. smithii* was observed to be 832 kyr (95% highest posterior density
943 interval 715K-971 kyr; Figure 3B). This directly precedes the split of Neandertals and
944 early anatomically modern humans, assuming this split occurred as early as 700,000
945 years ago as molecular dating currently suggests[86].
946
947 **16S rRNA amplicon comparisons to shotgun sequencing**
948 Rapid and inexpensive amplicon sequencing has been applied to ancient
949 metagenomes[1,31], despite the fact that amplicon libraries are biased by amplicon
950 selection, construction, and fragment length (reviewed in [87]). As a result of DNA
951 breakdown over time (taphonomy), ancient amplicon libraries may suffer from
952 additional bias introduced from short DNA fragment lengths. To determine the
953 efficacy of 16S rRNA amplicon sequencing on ancient samples, the 16S rRNA and
954 shotgun sequencing (>10 million reads per sample) data sets were compared for five
955 samples: chimpanzee, El Sidrón 1, El Sidrón 2, Spy 1, Spy 2, and modern human
956 C10. We first compared the 16S rRNA amplicon libraries to shotgun metagenomes
957 constructed from the same ancient and modern samples (Figure 1, S22, and S25). This
958 revealed drastically different microbial community structures between the two
959 approaches (Table S14). Differences between the samples were not limited to specific
960 bacterial phyla (Figures S24 and 25), as Gram-positive, Gram-negative, and archaeal

961 phyla were all impacted (Table S14) despite different outer cell wall structures. In
962 fact, UPGMA clustering of Euclidean distances revealed that in most cases filtered
963 and unfiltered ancient 16S rRNA libraries were more similar to themselves than to the
964 shotgun data produced from their respective samples (Figure 1). In contrast, data sets
965 generated from a modern specimen clustered together and were not subject to the
966 same level of bias observed in ancient samples (Figure 1; Figure S25). There are two
967 key issues that likely contribute to this.

968 First, exogenous microbial DNA from the environment has been shown to
969 dominate porous ancient samples, such as bone. In addition, bacterial DNA can be
970 regularly isolated from laboratory reagents[34], likely overpowering the initial signal
971 from endogenous oral microorganisms. To assess the impacts of exogenous DNA on
972 these differences, data sets with the environmental and laboratory microorganisms
973 removed were compared to unfiltered data (Figure 1). While this did not correct the
974 differences between the shotgun and amplicon sequencing approaches (Figure S24), it
975 did enhance the endogenous signal from the historic specimen (chimpanzee) and one
976 ancient sample (El Sidrón 2) (Figure 1).

977 Second, the damage and fragmentation of ancient DNA is also likely to limit
978 species detection and identification in amplicon libraries. 16S rRNA reads in ancient
979 metagenomes averaged 74.7 bp (Table S6) – well below the size required for robust
980 taxa identification (100 bp) or detection with gold-standard 16S primer sets (269
981 bp)[88]. However, larger 16S sequences (up to 181 bp) were identified in the shotgun
982 data set, and their community structure was comparable to the profile from the whole
983 shotgun data set (Figure S13), suggesting that shorter 16S rRNA amplicon fragments
984 or hybridization enrichment may provide more accurate pictures of ancient
985 community diversity. In addition, UPGMA clustering filtered 16S rRNA amplicon

libraries with 25 additional amplicon libraries generated from samples by Adler *et al*[1] (n=49) (Table S1) using unweighted UniFrac[89] was able to resolve some differences between ancient samples (ancient agriculturalists from hunter-gatherers; ancient agriculturalists from modern humans; Figure S7), suggesting that some longer single stranded 16S rRNA fragments are stable in dental calculus and that increased sample size of amplicon data may provide useful information in some cases, likely with less ancient material.

Earlier studies have hypothesized that taphonomic biases of ancient bacterial communities may occur different in select phyla, *i.e.* those with weaker outer cell wall structures compared to those with robust peptidoglycan matrices[1] or in specific taxa with smaller 16S rRNA sizes in specific regions (V3)[90]. To determine if differences between 16S amplicon and shotgun data sets were limited to select phyla, filtered and unfiltered data sets were directly compared (Figures S24-25). Each ancient sample had biases in certain phyla, although only biases in archaeal phyla appear to be shared across multiple ancient specimens (Table S14), which is likely a result of bias introduced during 16S amplicon library construction using conserved bacterial primer sets[90]. Overall, Gram-positive and Gram-negative phyla were equally impacted, and no particular phyla appeared to be systemically biased in all ancient specimens. Together, these observations highlight significant issues with solely utilizing 16S rRNA amplification for the analysis of ancient metagenomes and indicate that further sequencing of larger data sets is necessary to understand the taphonomic processes at work on ancient bacterial communities.

1008 **References**

1009 1. Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in
1010 oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat.*
1011 *Genet.* **45,** 450–455 (2013).
1012 2. Price, T. D. *et al.* in *Fundberichte aus Baden-Württemberg* (ed. Funda, D. T.) 23–
1013 58 (Kommissionsverlag Konrad Theiss Verlag, 2003).
1014 3. Semal, P. *et al.* New data on the late Neandertals: Direct dating of the Belgian
1015 Spy fossils. *Am. J. Phys. Anthropol.* **138,** 421–428 (2009).
1016 4. Lalueza-Fox, C. *et al.* Genetic evidence for patrilocal mating behavior among
1017 Neandertal groups. *Proc. Natl. Acad. Sci.* **108,** 250–253 (2011).
1018 5. Dean, M. C. *et al.* Longstanding dental pathology in Neandertals from El Sidrón
1019 (Asturias, Spain) with a probable familial basis. *J. Hum. Evol.* **64,** 678–686
1020 (2013).
1021 6. Manzi, G. & Passarello, P. At the Archaic/Modern Boundary of the Genus Homo:
1022 The Neandertals From Grotta Breuil. *Curr. Anthropol.* **36,** 355–366 (1995).
1023 7. Buckley, S., Usai, D., Jakob, T., Radini, A. & Hardy, K. Dental Calculus Reveals
1024 Unique Insights into Food Items, Cooking and Plant Processing in Prehistoric
1025 Central Sudan. *PLoS ONE* **9,** e100808 (2014).
1026 8. Arkell, A. J. *Shaheinab, an Account of the Excavation of a Neolithic Occupation*
1027 *Site Carried Out for the Sudan Antiquities Service in 1949-50, by A. J. Arkell,...*
1028 (Oxford University Press, 1953).
1029 9. The Lech Krzyzaniak Excavations in the Sudan, Kadero, Studies in Afric.
1030 Available at: http://www.archeobooks.com/products/the-lech-krzyzaniak-
1031 excavations-in-the-sudan-kadero-studies-in-african-archaeology-vol-10.
1032 (Accessed: 9th October 2015)

10. Madella, M., García-Granero, J. J., Out, W. A., Ryan, P. & Usai, D. Microbotanical Evidence of Domestic Cereals in Africa 7000 Years Ago. *PLoS ONE* **9,** e110177 (2014).
11. Morris, A. G. *A Master Catalogue: Holocene Human Skeletons from South Africa*. (Witwatersrand Univ Pr, 1993).
12. Mitchell, P. *The Archaeology of Southern Africa*. (Cambridge University Press, 2002).
13. Mountain, A. *The First People of the Cape: A Look at Their History and the Impact of Colonialism on the Cape's Indigenous People*. (New Africa Books, 2003).
14. Sealy, J. C., Patrick, M. K., Morris, A. G. & Alder, D. Diet and dental caries among later stone age inhabitants of the Cape Province, South Africa. *Am. J. Phys. Anthropol.* **88,** 123–134 (1992).
15. Reenen, J. F. V. Dental Features of a Low-Caries Primitive Population. *J. Dent. Res.* **45,** 703–713 (1966).
16. Morris, A. G. *et al.* Later Stone Age burials from the Western Cape Province, South Africa Part 1: Voelvlei. *South. Afr. Field Archaeol.* **13 & 14,** 19–26
17. Cronin, M. Radiocarbon dates for the Early Iron Age in Transkei. *South Afr. J. Sci.* **78,** 38–39 (1982).
18. Morris, A. *The Skeletons of Contact: Protohistoric burials from the lower Orange River Valley*. (Johannesburg: Witwatersrand University Press, 1992).
19. Wiegelmann, H. j T. and G. *Der Wandel der Nahrungsgewohnheiten unter dem Einfluß der Industrialisierung. (Studien zum Wandel von Gesellschaft und Bildung im 19. Jahrhundert)*. (Göttingen: Vandenhoek & Ruprecht).

20. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Adler CJ, Richards SM, Der Sarkissian C, Ganslmeier R, Friederich S, Dresely V, van Oven M, Kenyon R, Van der Hoek M, Korlach J, Luong K, Ho SYW, Quintana-Murci L, Behar DM, Meller H, Alt KW, Cooper A & The Genographic Project. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* **4**:1764.
21. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6,** 1621–1624 (2012).
22. Cooper, A. & Poinar, H. N. Ancient DNA: Do It Right or Not at All. *Science* **289,** 1139–1139 (2000).
23. Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V. & Polz, M. F. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Appl. Environ. Microbiol.* **71,** 8966–8969 (2005).
24. Gonzalez, J. M., Portillo, M. C., Belda-Ferre, P. & Mira, A. Amplification by PCR Artificially Reduces the Proportion of the Rare Biosphere in Microbial Communities. *PLoS ONE* **7,** e29973 (2012).
25. Aronesty, E. *Command-line tools for processing biological sequencing data*. (2011).
26. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
27. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. 335–336 (2010).
28. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. 2460–2461 (2010).

29. DeSantis, T. Z. *et al.* Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).
30. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Na?ve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
31. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* **46,** 336–344 (2014).
32. Rougier, H. *Spy Cave: 125 Years of Multidisciplinary Research at the Betche Aux Rotches (Jemeppe-sur-Sambre, Province of Namur, Belgium)*. (Société Royale Belge d'Anthropologie et de Préhistoire, 2012).
33. Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. & Pääbo, S. Ancient DNA. *Nat. Rev. Genet.* **2,** 353–359 (2001).
34. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12,** (2014).
35. Schmieder, R. & Edwards, R. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* **6,** e17288 (2011).
36. Ramirez, K. S., Lauber, C. L. & Fierer, N. Microbial consumption and production of volatile organic compounds at the soil-litter interface. *Biogeochemistry* **99,** 97–107 (2010).
37. Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* (2013). doi:10.1038/ng.2536

38. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database J. Biol. Databases Curation* **2010,** baq013 (2010).
39. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010,** pdb prot5448 (2010).
40. Herbig, A. *et al.* MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* 50559 (2016). doi:10.1101/050559
41. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Meth* **12,** 59–60 (2015).
42. Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. 1552–1560 (2011).
43. Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinforma. Oxf. Engl.* **14,** 68–73 (1998).
44. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12,** R60 (2011).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
46. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13,** 178 (2012).
47. Kircher, M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* **840,** 197–228 (2012).

48. Ginolhac, A., Rasmussen, M., Gilbert, M. T., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27,** 2153–5 (2011).
49. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. 1647–1649 (2012).
50. Krzywinski, M. I. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* (2009). doi:10.1101/gr.092759.109
51. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5,** e11147 (2010).
52. Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* **56,** 564–577 (2007).
53. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. 2688–2690 (2006).
54. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–90 (2006).
55. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
56. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25,** 1422–1423 (2009).
57. Longo, M. S., O'Neill, M. J. & O'Neill, R. J. Abundant Human DNA Contamination Identified in Non-Primate Genome Databases. *PLOS ONE* **6,** e16410 (2011).

58. Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N. C. & Pati, A. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand. Genomic Sci.* **10,** 18 (2015).

59. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **8,** 761–763 (2011).

60. Hewitt, K. M. *et al.* Bacterial Diversity in Two Neonatal Intensive Care Units (NICUs). *PLoS ONE* **8,** e54703 (2013).

61. Allentoft, M. E. *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B Biol. Sci.* rspb20121745 (2012). doi:10.1098/rspb.2012.1745

62. R Development Core Team. *R: A language and envrionment for statistical computing, reference index version 2.x.x.* (R Foundation for Statistical Computing).

63. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4,** 102–112 (2006).

64. Bollongino, R. *et al.* 2000 Years of Parallel Societies in Stone Age Central Europe. *Science* **342,** 479–481 (2013).

65. Gumiński, W. *The peat-bog site Dudka, Masurian Lakeland: an example of conservative economy.* (Sheffield Academic Press, 1998).

66. Poinar, H. N. *et al.* Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311,** 392–394 (2006).

67. Frampton, M. & Houlston, R. Generation of Artificial FASTQ Files to Evaluate the Performance of Next-Generation Sequencing Pipelines. *PLoS ONE* **7,** e49110 (2012).

68. Meyer, F. *et al.* The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. 386 (2008).
69. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9,** 811–814 (2012).
70. Lindgreen, S., Adair, K. L. & Gardner, P. P. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci. Rep.* **6,** 19233 (2016).
71. Belda-Ferre, P. *et al.* The oral metagenome in health and disease. *ISME J.* **6,** 46–56 (2012).
72. Lazarevic, V. *et al.* Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J. Microbiol. Methods* **79,** 266–271 (2009).
73. Cameron, S. J. S., Huws, S. A., Hegarty, M. J., Smith, D. P. M. & Mur, L. A. J. The human salivary microbiome exhibits temporal stability in bacterial diversity. *FEMS Microbiol. Ecol.* **91,** fiv091 (2015).
74. Moeller, A. H. *et al.* Rapid changes in the gut microbiome during human evolution. *Proc. Natl. Acad. Sci.* **111,** 16431–16435 (2014).
75. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
76. Briggs, A. W. *et al.* Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* **325,** 318–321 (2009).
77. Schuenemann, V. J. *et al.* Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science* **341,** 179–183 (2013).
78. Ozga, A. T. *et al.* Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* **160,** 220–228 (2016).

79. Brockmeier, S. L. & Register, K. B. Expression of the dermonecrotic toxin by Bordetella bronchiseptica is not necessary for predisposing to infection with toxigenic Pasteurella multocida. *Vet Microbiol* **125,** 284–9 (2007).
80. Parkhill, J. *et al.* Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nat Genet* **35,** 32–40 (2003).
81. Scarre, C. *The Human Past: World Prehistory and the Development of Human Societies*. (Thames & Hudson, 2013).
82. Bocquet-Appel, J.-P. & Bar-Yosef, O. *The Neolithic Demographic Transition and its Consequences*. (Springer Science & Business Media, 2008).
83. van der Zee, A., Mooi, F., Van Embden, J. & Musser, J. Molecular evolution and host adaptation of Bordetella spp.: phylogenetic analysis using multilocus enzyme electrophoresis and typing with three insertion sequences. *J. Bacteriol.* **179,** 6609–6617 (1997).
84. Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. 506–510 (2011).
85. Duchêne, S., Geoghegan, J. L., Holmes, E. C. & Ho, S. Y. Estimating evolutionary rates using time-structured data: a general comparison of phylogenetic methods. *Bioinformatics* btw421 (2016). doi:10.1093/bioinformatics/btw421
86. Stringer, C. The origin and evolution of Homo sapiens. *Phil Trans R Soc B* **371,** 20150237 (2016).
87. Sharpton, T. J. An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* **5,** (2014).

88. Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D. & Knight, R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* **35,** e120 (2007).

89. Lozupone, C. & Knight, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. 8228–8235 (2005).

90. Ziesemer, K. A. *et al.* Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Sci. Rep.* **5,** 16498 (2015).

Table S1 – Sample names, descriptions and archaeological information associated with each sample used within this study. If calculus samples were analyzed in a previous study, the reference is given, and the data obtained from each sample is also noted.

ACAD Number	Description	Figure Abbreviation	Host Species	Culture or Group	Archaeological Site Location	Museum Sample Number	Museum	Date(yBP)	16S or Shotgun Sequencing	Calculus Sample Reference
12055	Neandertal – Spy I	Spy 1	Homo neandertalensis	Paleolithic	Spy Cave, Belgium	Spy I (94a)	Royal Belgian Institute of Natural Sciences, Brussels, Belgium	35,810+260	16S/Shotgun	(this study)
14017	Neandertal – Spy II	Spy 2	Homo neandertalensis	Paleolithic	Spy Cave, Belgium	Spy II (92b)	Royal Belgian Institute of Natural Sciences, Brussels, Belgium	36,350+310	16S/Shotgun	(this study)
12056	Neandertal - El Sidrón1	ElSidron1	Homo neandertalensis	Paleolithic	El Sidrón Cave, Spain	El Sidrón 1 SD 1427C	Museo Nacional de Ciencias Naturales, Madrid	~49,000yBP	16S/Shotgun	(this study)
12057	Neandertal - El Sidrón2	ElSidron2	Homo neandertalensis	Paleolithic	El Sidrón Cave, Spain	El Sidrón 6 SD 1604	Museo Nacional de Ciencias Naturales, Madrid	~49,000yBP	16S/Shotgun	(this study)
8911	Neandertal - Brueil Grotta	N/A	Homo neandertalensis	Paleolithic	Brueil Grotta, Italy	8911	University of Florence, Italy	nd	nd	(this study)
12014	Mesolithic - Poland 1	EuroHG1	Homo sapiens	Hunter-gatherer	Dudka, Poland	nd	University of Warsaw, Poland	~7550	16S/Shotgun	1
12017	Mesolithic - Poland 2	EuroHG2	Homo sapiens	Hunter-gatherer	Dudka, Poland	VI-7 A	University of Warsaw, Poland	~7550	16S/Shotgun	1
12826	LBK - Stuttgart, Germany 1	LBKS1	Homo sapiens	Neolithic	Stuttgart-Mühlhausen I, Germany	762	Landesdenkmalpflege Baden-Württemberg, University of Tubingen, Germany	~7440	16S/Shotgun	(this study)
12829	LBK - Stuttgart, Germany 2	LBKS2	Homo sapiens	Neolithic	Stuttgart-Mühlhausen I, Germany	765	Landesdenkmalpflege Baden-Württemberg, University of Tubingen, Germany	~7440	16S/Shotgun	(this study)
12824	LBK - Stuttgart, Germany 3	LBKS3	Homo sapiens	Neolithic	Stuttgart-Mühlhausen I, Germany	760	Landesdenkmalpflege Baden-Württemberg, University of Tubingen, Germany	~7440	16S/Shotgun	(this study)
12830	LBK - Stuttgart, Germany 4	LBKS4	Homo sapiens	Neolithic	Stuttgart-Mühlhausen I, Germany	766	Landesdenkmalpflege Baden-Württemberg, University of Tubingen, Germany	~7440	16S	(this study)
8240	LBK - Hallberstadt, Germany 1	LBKH	Homo sapiens	Neolithic	LBK - Halberstadt-Sonntagsfeld 1	HAL25b	Landesmuseum Sachsen-Anhalt, Germany	~7440	16S	1
8277	LBK - Hallberstadt, Germany 2	LBKH	Homo sapiens	Neolithic	LBK - Halberstadt-Sonntagsfeld 2	HAL39a	Landesmuseum Sachsen-Anhalt, Germany	~7440	16S	1
11105	Early Neolithic - Sudan 1	AfrSud1	Homo sapiens	African Early Neolithic	Al Khiday, Sudan	16D4 grave93	Durham University, England	~5,000yBP	16S/Shotgun	(this study)
11106	Early Neolithic - Sudan 2	AfrSud2	Homo sapiens	African Early Neolithic	Al Khiday, Sudan	16D4 grave96	Durham University, England	~5,000yBP	16S/Shotgun	(this study)
4331	BellBeaker - Germany	BB	Homo sapiens	Bell Beaker	BellBeaker - Quedlinburg XII	Befund 967	Landesmuseum Sachsen-Anhalt, Germany	~4450	16S	1
9436	LNBA - Germany	LN/BA	Homo sapiens	Late Neolithic, Bell Beaker	LNBA - Benzingerode-Heimburg	BZH17	Landesmuseum Sachsen-Anhalt, Germany	~4150	16S	1
8890	Bronze Age - England 1	BA1	Homo sapiens	Bronze Age	Yorkshire, England	T82GF	University of Aberdeen, Scotland	~4150	16S	1
8891	Bronze Age - England 2	BA2	Homo sapiens	Bronze Age	Yorkshire, England	14Barrow163	University of Aberdeen, Scotland	~4150	16S	1
13207	Pre-pastoralist Period - South Africa 2	AfrSF1	Homo sapiens	African LSA	Near Cape Town, South Africa	UCT 169	Department of Human Biology of the University of Cape Town, South Africa	2320+/-50yrsBP	16S	(this study)
13208	Pre-pastoralist Period - South Africa 3	AfrSF2	Homo sapiens	African LSA	Near Cape Town, South Africa	UCT 373	Department of Human Biology of the University of Cape Town, South Africa	3835+/-50yrsBP	16S	(this study)
13209	Pre-pastoralist Period - South Africa 4	AfrSF3	Homo sapiens	African LSA	Near Cape Town, South Africa	UCT 386	Department of Human Biology of the University of Cape Town, South Africa	2000+/-50yrsBP	16S/Shotgun	(this study)
13210	Pre-pastoralist Period - South Africa 5	AfrSF4	Homo sapiens	African LSA	Near Cape Town, South Africa	UCT 421	Department of Human Biology of the University of Cape Town, South Africa	2895+/-45BP	16S/Shotgun	(this study)
13211	Pre-pastoralist Period - South Africa 6	AfrSF5	Homo sapiens	African LSA	Near Cape Town, South Africa	UCT 427	Department of Human Biology of the University of Cape Town, South Africa	2670+/-80yrsBP	16S	(this study)
Warinner_B61	Medieval - Germany	War_B61	Homo sapiens	Dalheim	Dalheim, Germany	B61	University of Zürich's Institute of Anatomy, Switzerland	c.950–1200CE	Shotgun	31
Warinner_G12	Medieval - Germany	War_G12	Homo sapiens	Dalheim	Dalheim, Germany	G12	University of Zürich's Institute of Anatomy, Switzerland	c.950–1200CE	Shotgun	31
8333	Early Rural Medieval - England 1	RM1	Homo sapiens	Raunds Furnells	Northamptonshire, England	R5287	University of Aberdeen, Scotland	~1100	16S	1
8335	Early Rural Medieval - England 2	RM2	Homo sapiens	Raunds Furnells	Northamptonshire, England	R5252	University of Aberdeen, Scotland	~1100	16S	1
8336	Early Rural Medieval - England 3	RM3	Homo sapiens	Raunds Furnells	Northamptonshire, England	R5008	University of Aberdeen, Scotland	~1100	16S	1
8869	Early Rural Medieval - England 4	RM4	Homo sapiens	Raunds Furnells	Northamptonshire, England	R5229	University of Aberdeen, Scotland	~1100	16S	1
8874	Late Rural Medieval - St Helens, England 11	RMS1	Homo sapiens	St. Helens on the Walls	York, England	5241	University of Aberdeen, Scotland	~1100	16S	1
8330	Late Urban Medieval - England 1	UM1	Homo sapiens	Jewbury	York, England	J2303	Jewbury Cemetery, England	~750	16S	1
8332	Late Urban Medieval - England 2	UM2	Homo sapiens	Jewbury	York, England	J1336	Jewbury Cemetery, England	~750	16S	1
8477	Late Urban Medieval - England 3	UM3	Homo sapiens	Jewbury	York, England	2357	Jewbury Cemetery, England	~750	16S	1
8482	Late Urban Medieval - England 5	UM4	Homo sapiens	Jewbury	York, England	2358	Jewbury Cemetery, England	~750	16S	1
8812	Late Urban Medieval - England 1 (QG)	UM5	Homo sapiens	Jewbury	York, England	J2360	Jewbury Cemetery, England	~750	16S/Shotgun	(this study)
8824	Late Urban Medieval - England 2 (QG)	UM6	Homo sapiens	Jewbury	York, England	J2454	Jewbury Cemetery, England	~750	16S/Shotgun	(this study)
13213	Pastoralist Period - South Africa 1	AfrPP1	Homo sapiens	African Pastoralist Period	Near Cape Town, South African	UCT 582	Department of Human Biology of the University of Cape Town, South Africa	740+/-40yrsBP	16S/Shotgun	(this study)
13204	Pastoralist Period - South Africa 2	AfrPP2	Homo sapiens	African Pastoralist Period	Near Cape Town, South African	UCT 67	Department of Human Biology of the University of Cape Town, South Africa	570+/-45yrsBP	16S/Shotgun	(this study)
13206	Pastoralist Period - South Africa 3	AfrPP3	Homo sapiens	African Pastoralist Period	Near Cape Town, South African	UCT 157	Department of Human Biology of the University of Cape Town, South Africa	587+/-28yrsBP	16S	(this study)
13227	Post-Industrial Revolution - Germany 1	IndRev1	Homo sapiens	Post-Industrial Revolution	Hettstedt Gymnasium, Germany	23	Johannes Gutenberg University, Mainz, Germany	1860-1865	16S	(this study)
13230	Post-Industrial Revolution - Germany 2	IndRev2	Homo sapiens	Post-Industrial Revolution	Hettstedt Gymnasium, Germany	43	Johannes Gutenberg University, Mainz, Germany	1860-1865	16S/Shotgun	(this study)
13232	Post-Industrial Revolution - Germany 3	IndRev3	Homo sapiens	Post-Industrial Revolution	Hettstedt Gymnasium, Germany	85	Johannes Gutenberg University, Mainz, Germany	1860-1865	16S/Shotgun	(this study)
P2	Modern Plaque 2	Modern P2	Homo sapiens	Modern Plaque	Adelaide, Australia	P2	University of Adelaide Dental School, Australia	0	16S	1
P6	Modern Plaque 6	Modern P6	Homo sapiens	Modern Plaque	Adelaide, Australia	P6	University of Adelaide Dental School, Australia	0	16S	1
P7	Modern Plaque 7	Modern P7	Homo sapiens	Modern Plaque	Adelaide, Australia	P7	University of Adelaide Dental School, Australia	0	16S	1
P8	Modern Plaque 8	Modern P8	Homo sapiens	Modern Plaque	Adelaide, Australia	P8	University of Adelaide Dental School, Australia	0	16S	1
P10	Modern Plaque 10	Modern P10	Homo sapiens	Modern Plaque	Adelaide, Australia	P10	University of Adelaide Dental School, Australia	0	16S	1
C5	Modern Calculus 5	Modern C5	Homo sapiens	Modern Calculus	Adelaide, Australia	C5	University of Adelaide Dental School, Australia	0	16S	1
C8	Modern Calculus 6	Modern C6	Homo sapiens	Modern Calculus	Adelaide, Australia	C6	University of Adelaide Dental School, Australia	0	16S	1
C7	Modern Calculus 7	Modern C7	Homo sapiens	Modern Calculus	Adelaide, Australia	C7	University of Adelaide Dental School, Australia	0	16S	1
C10	Modern Calculus 10	Modern C10	Homo sapiens	Modern Calculus	Adelaide, Australia	C10	University of Adelaide Dental School, Australia	0	16S/Shotgun	1
12873	Chimpanzee	Chimp	Pan troglodytes	Primate, wild caught	Gola Forest, Sierra Leone	810	Odontological Collection, The Royal College of Surgeons, England	~70yBP	16S/Shotgun	(this study)
12878	Macaque	N/A	Macaca mulatta	Primate, captive	nd	814	Odontological Collection, The Royal College of Surgeons, England	nd	16S	(this study)
12887	Baboon	N/A	Papio anubis	Primate, captive	nd	815	Odontological Collection, The Royal College of Surgeons, England	nd	16S	(this study)
8483	Tooth - England	N/A	Environmental	Tooth	York, England	2665	Jewbury Cemetery, England	~750	16S	1

Figure S1 – Photographs of representative dental calculus samples prior to decontamination. A scale bar of 0.5 cm is included for reference.

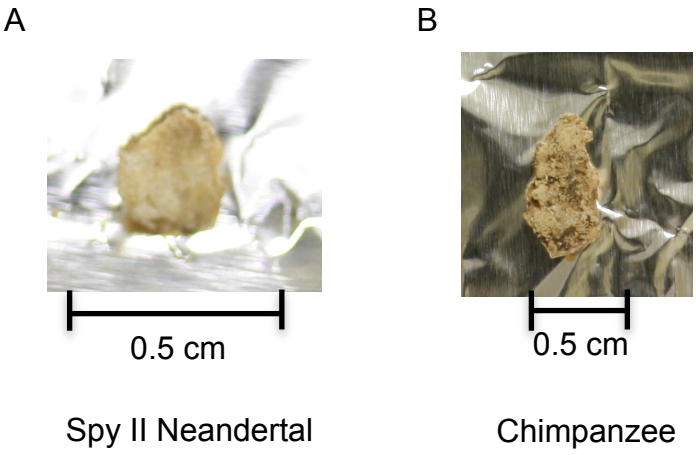


Figure S2 – OTUs post-filtering to test the impact of different extraction methods on retrieval of bacterial communities and investigate PCR bias during 16S metagenomic studies of ancient DNA on a Medieval (Jewbury), Neandertal (Neandertal Spy I), ancient hunter-gatherer (Mesolithic Poland) or a chimpanzee (Chimpanzee) sample. A comparison between Late Medieval samples extract using two methods was examined, and three different cultures using either 37 cycles or 43 cycles of PCR under otherwise similar conditions was also examined.

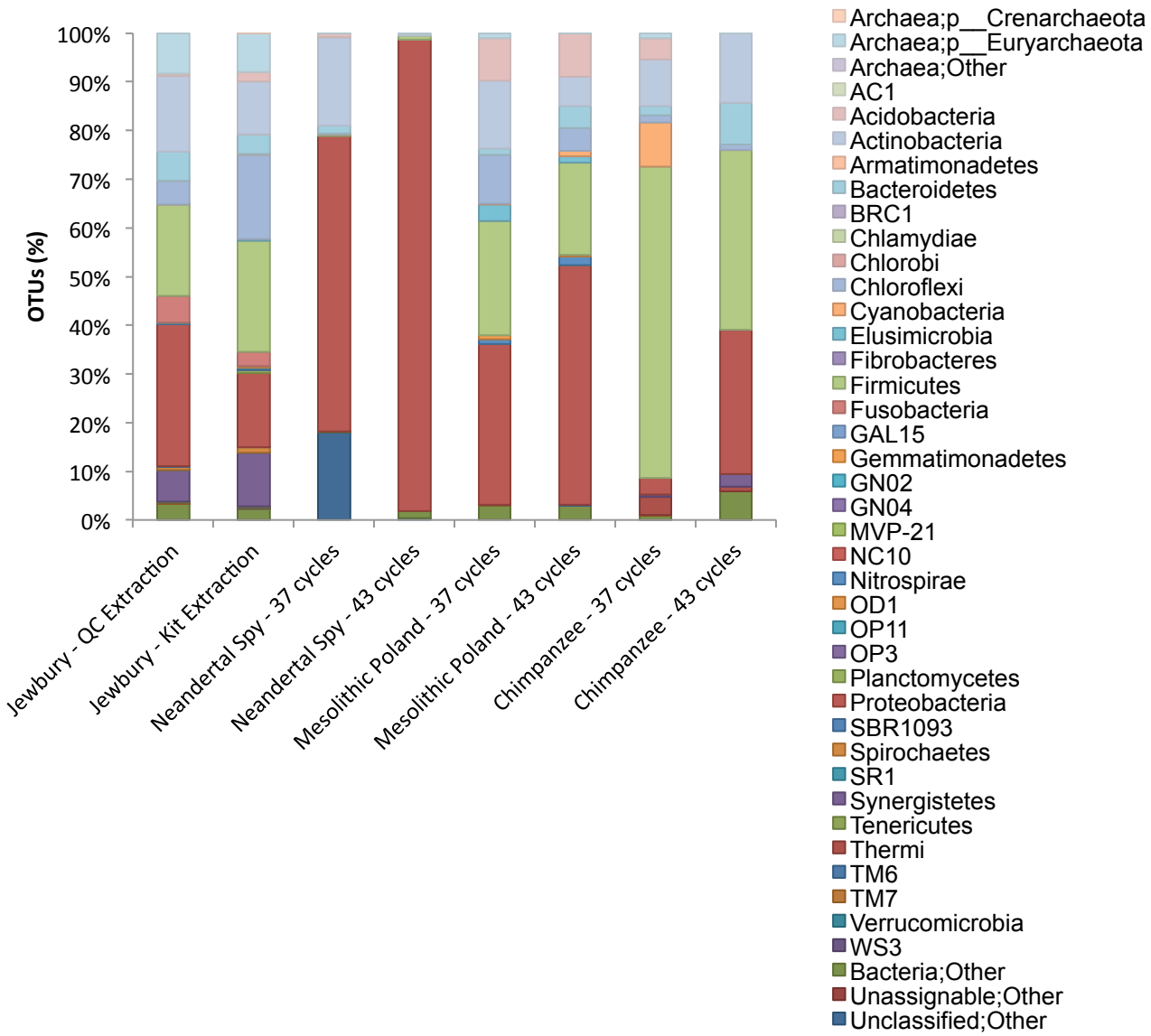


Table S2 – The number of OTUs is monitored throughout the filtering process. The total number of reads removed and the percent of total reads discarded is displayed for each filtering step. Samples that did not meet the OTU threshold and were removed from downstream analysis (red) (SI Figure 4), while other samples were removed due to lack of accurate provenance or excluded because they were included as methodological replicates, but were retained in the SI figures and analysis for reference (orange).

SampleName	Cleaned/Raw Reads	Reads Post EBC Filtering	% Reads Discarded	Reads after Soil Filtering	% Reads Discarded	Reads after known contam filtering	% Reads Discarded	Total Reads Filtered	Total Reads (Oral)	Total % Reads Filtered
LBKS3	1449685	286341	80.2	281959.0	1.5	281308	0.2	80.6	162454.0	88.8
Chimp	791712	42505	94.6	42445.0	0.1	41182	3.0	94.8	32498.0	95.9
AfrPP1	778898	69025	91.1	69000.0	0.0	60480.0	12.3	92.2	49675.0	93.6
AfrSF5	596701	48042	91.9	48025.0	0.0	17586	63.4	97.1	17563.0	97.1
AfrSF4	474139	4783	99.0	4767.0	0.3	3732	21.7	99.2	3705.0	99.2
Neandertal – Spy I	462176	7778	98.3	7757.0	0.3	6670	14.0	98.6	5398.0	98.8
EuroHG2	419370	65876	84.3	64625.0	1.9	62309.0	3.6	85.1	39073.0	90.7
EuroHG1	409017	39170	90.4	39162.0	0.0	28749	26.6	93.0	26047.0	93.6
AfriPP2	380802	32062	91.6	31503.0	1.7	31336	0.5	91.8	22293.0	94.1
Modern C7	366603	45541	87.6	44648.0	2.0	44643	0.0	87.8	43291.0	88.2
IndRev1	346510	45080	87.0	44542.0	1.2	44518	0.1	87.2	32856.0	90.5
Modern C5	291597	46174	84.2	45824.0	0.8	45815	0.0	84.3	44444.0	84.8
IndRev2	284655	47278	83.4	46282.0	2.1	46260	0.0	83.7	42171.0	85.2
LBKS1	284453	37351	86.9	37134.0	0.6	35198	5.2	87.6	23853.0	91.6
LM5	283204	38482	86.4	38057.0	1.1	38046	0.0	86.6	18538.0	93.5
LM3	272488	19536	92.8	19358.0	0.9	19352	0.0	92.9	12195.0	95.5
BB	266414	11391	95.7	11233.0	1.4	11192	0.4	95.8	7400.0	97.2
Modern C6	261108	43404	83.4	42910.0	1.1	42903	0.0	83.6	42049.0	83.9
Modern C10	242510	29261	87.9	28922.0	1.2	28917	0.0	88.1	28248.0	88.4
Modern P2	239966	13416	94.4	13377.0	0.3	13371	0.0	94.4	13145.0	94.5
LM1 (QG)	239925	53686	77.6	52803.0	1.6	52792	0.0	78.0	39935.0	83.4
EuroHG 1 (43)	237920	6438	97.3	6426.0	0.2	5729	10.8	97.6	5430.0	97.7
BA2	235263	10133	95.7	9987.0	1.4	9967	0.2	95.8	7013.0	97.0
Modern P8	233069	40187	82.8	40006.0	0.5	39999	0.0	82.8	39812.0	82.9
EM4	225985	15296	93.2	15191.0	0.7	15178	0.1	93.3	5392.0	97.6
LBKS2	217339	27787	87.2	27577.0	0.8	27326	0.9	87.4	19246.0	91.1
LBKS4	213285	26128	87.7	25858.0	1.0	25819	0.2	87.9	15204.0	92.9
AfrSF3	212040	33494	84.2	32842.0	1.9	32783	0.2	84.5	26569.0	87.5
EM3	211172	32321	84.7	32063.0	0.8	32042	0.1	84.8	14708.0	93.0
AfrSud2	203867	20104	90.1	20097.0	0.0	18365	8.6	91.0	6791.0	96.7
LBKH2	200689	7556	96.2	7439.0	1.5	7419	0.3	96.3	5940.0	97.0
IndRev3	200329	30921	84.6	30415.0	1.6	30410	0.0	84.8	27084.0	86.5
Modern P7	194582	20833	89.3	20679.0	0.7	20676	0.0	89.4	20441.0	89.5
EuroHG2 (43)	193317	22684	88.3	22651.0	0.1	19781	12.7	89.8	14655.0	92.4
BA1	184631	9952	94.6	9822.0	1.3	9788	0.3	94.7	7258.0	96.1
Modern P6	180900	14634	91.9	14586.0	0.3	14577	0.1	91.9	14367.0	92.1
Chimp (45)	180498	6537	96.4	6524.0	0.2	6173	5.4	96.6	5517.0	96.9
LNBA	177184	12549	92.9	12441.0	0.9	12304	1.1	93.1	5118.0	97.1
AfrSud1	172206	2255	98.7	2219.0	1.6	2018	9.1	98.8	1971.0	98.9
LBKH 1	166325	25925	84.4	25612.0	1.2	25530	0.3	84.7	8860.0	94.7
LM2 (QG)	158464	23800	85.0	23585.0	0.9	23582	0.0	85.1	17413.0	89.0
Neandertal – Spy II	158142	988	99.4	977.0	1.1	568	41.9	99.6	556.0	99.6
Modern P10	158090	36626	76.8	36539.0	0.2	36536	0.0	76.9	35926.0	77.3
LM11	154101	11424	92.6	11329.0	0.8	10998	2.9	92.9	7752.0	95.0
EM1	149276	14466	90.3	14239.0	1.6	14201	0.3	90.5	9821.0	93.4
LM1	146117	17670	87.9	17468.0	1.1	17463	0.0	88.0	11568.0	92.1
Macaque	143906	34161	76.3	33877.0	0.8	33786	0.3	76.5	24194.0	83.2
Tooth	130802	26018	80.1	25613.0	1.6	25602	0.0	80.4	21864.0	83.3
EM2	125471	28996	76.9	28703.0	1.0	28692.0	0.0	77.1	16273.0	87.0
Neandertal - El Sidron 2	120468	8428	93.0	8404.0	0.3	6761	19.6	94.4	6229.0	94.8
AfrSF2	112315	425	99.6	422.0	0.7	277	34.4	99.8	265.0	99.8
LM2	96682	11413	88.2	11171.0	2.1	11159	0.1	88.5	5925.0	93.9
Neandertal - El Sidron 1	3446	1467	57.4	1463.0	0.3	1411	3.6	59.1	487.0	85.9
Baboon	38	4	89.5	4.0	0.0	4	0.0	89.5	4.0	89.5
AfriPP3	26	1	96.2	1.0	0.0	1	0.0	96.2	1.0	96.2
AfrSF 6	17	7	58.8	7.0	0.0	7	0.0	58.8	4.0	76.5

Figure S3 – Unfiltered 16S rRNA OTUs identified by homology to the Greengenes database are plotted at the phyla level for the dental calculus and control samples used within this sample.

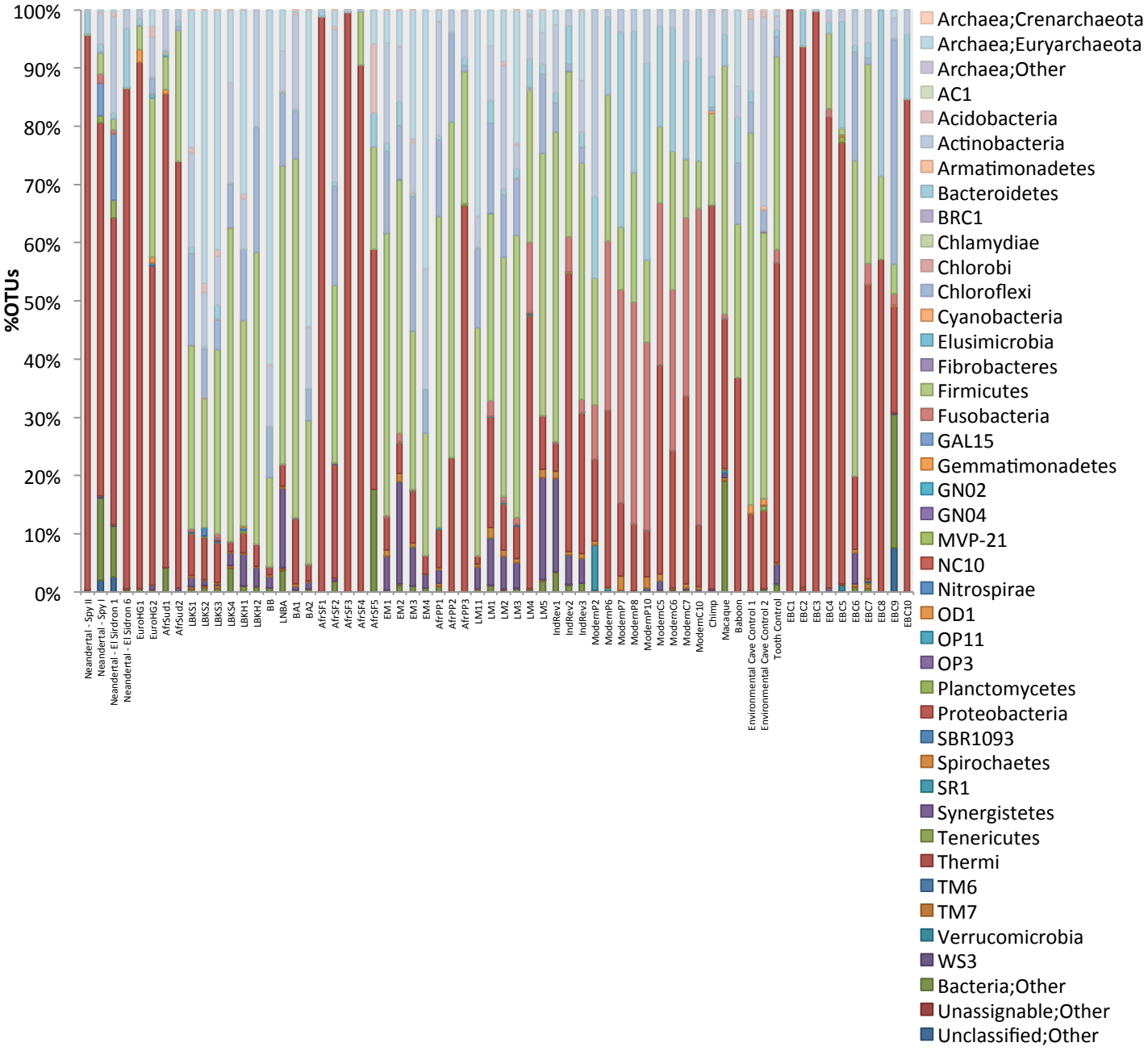


Figure S4 – The abundance of final 16S rRNA OTUs are presented for each sample, after environmental and laboratory contaminants OTUs have been removed.

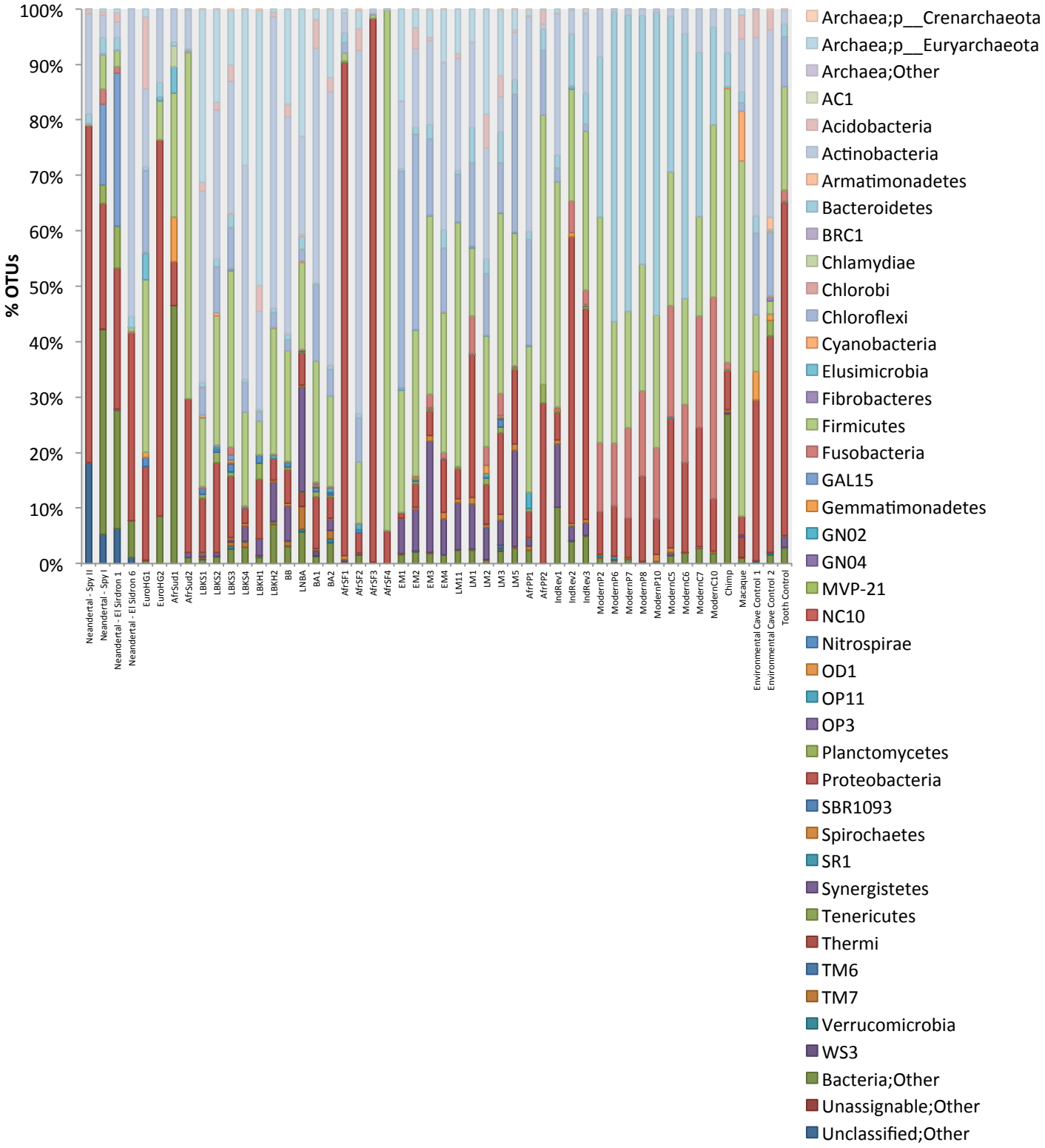


Figure S5 – Modern dental calculus samples sequenced in this study were compared to published data, including samples sequenced by Adler *et al.* 2013, data from the HOMD, and environmental controls, including soil, water, and inside archaeological teeth. Raw data was obtained, and the OTUs were picked alongside OTUs for the calculus analyzed in this study.

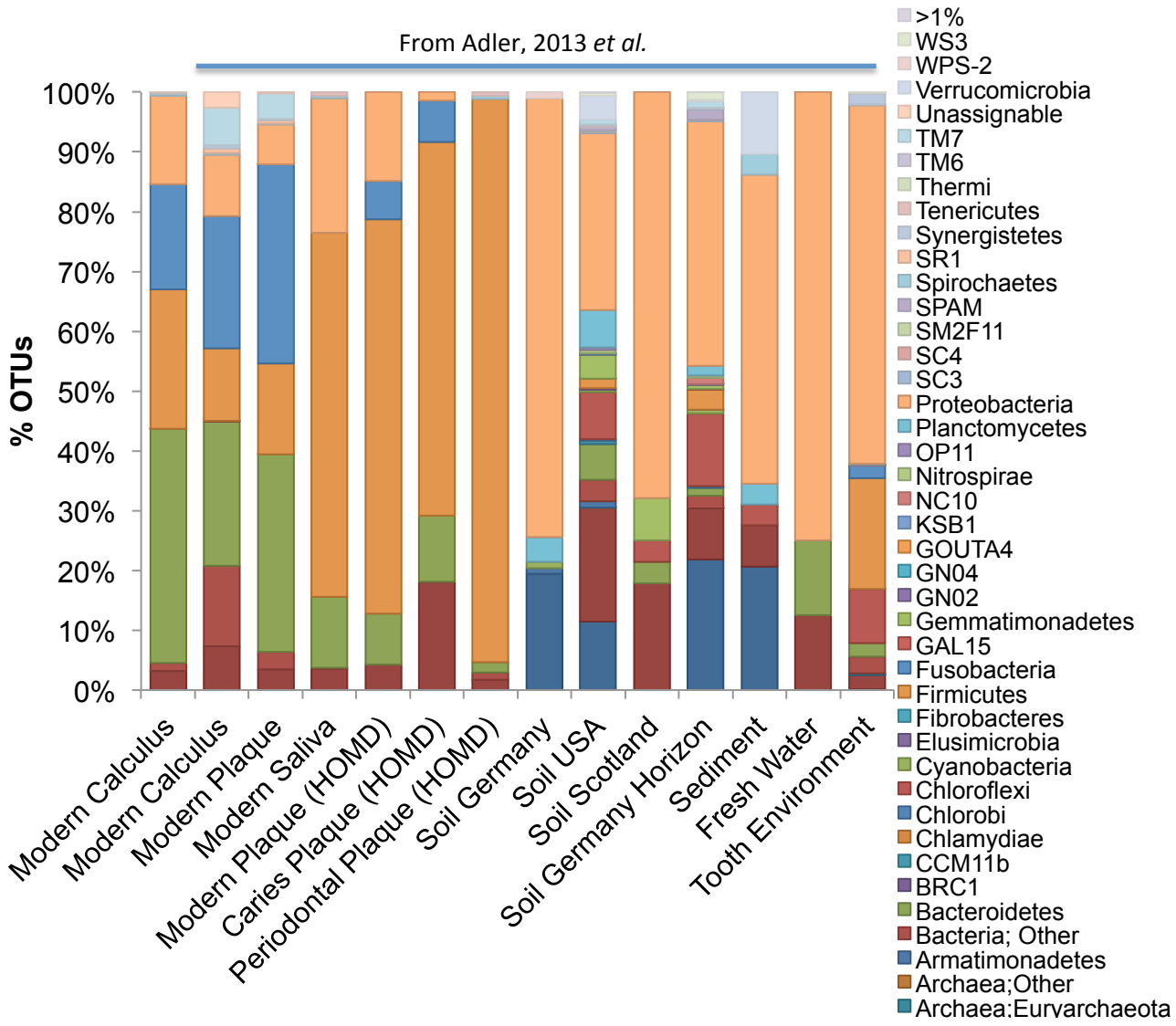


Figure S6 – SourceTracker take-one-out analysis was completed for all samples. Samples were grouped into time periods, the proportion of each taxa originating from each sample group was then inferred. "Other" represents summed proportions across non-oral microbial groups (non-oral human microbiome, air, and soil) and unknown classification. Groups have a minimum of two samples (Non-Human Primate group is removed in filtered analysis as filtering reduced sample number to one). Fig. A: from raw (unfiltered) OTU table, n = 54. Fig. B: from filtered OTU table, n = 42.

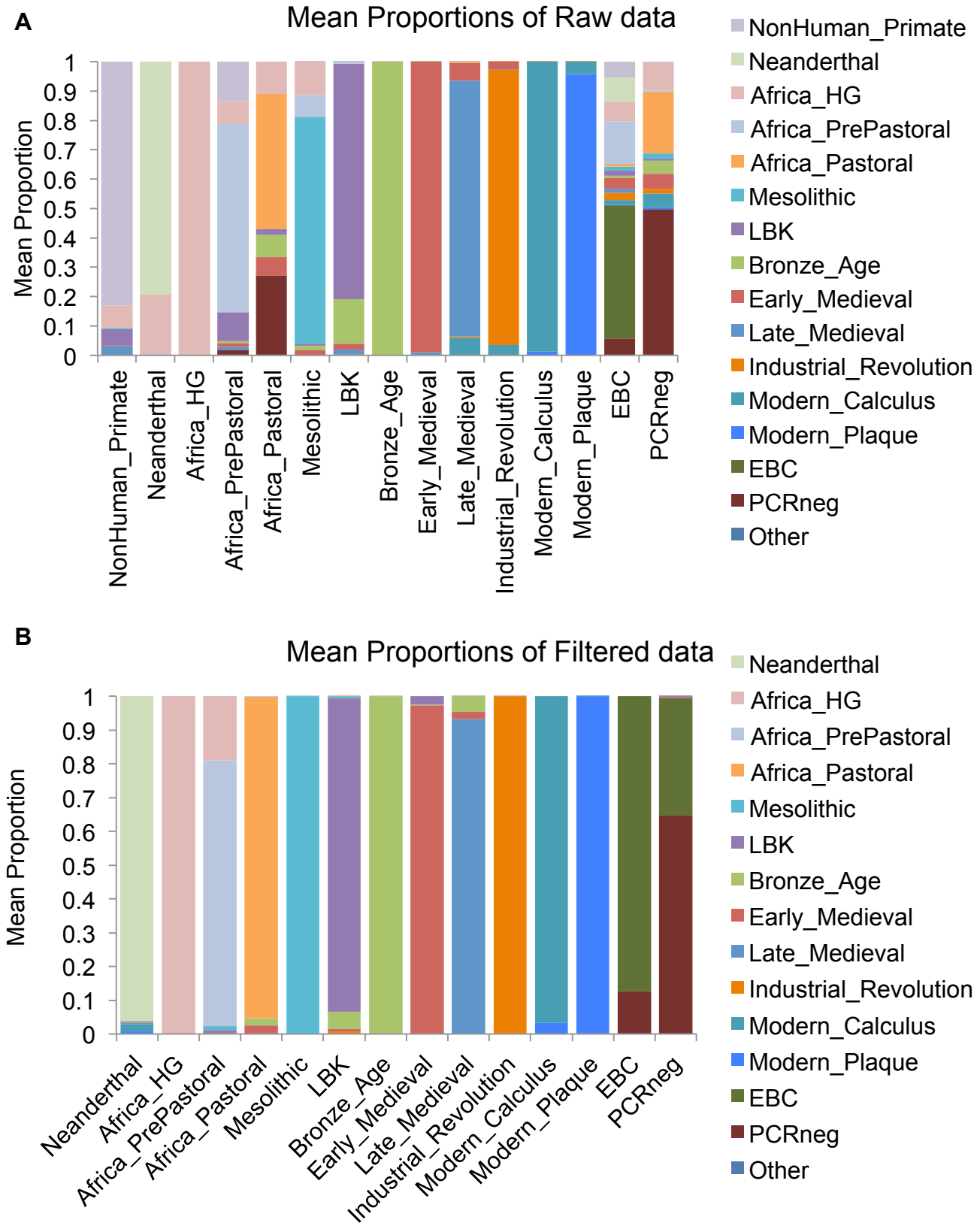


Figure S7 – As is consistent with ancient DNA metagenomic analysis, a presence/absence distance (Jaccard) were calculated for each OTU observed in the 99th percentile and clustered according to dissimilarity within each sample and amongst other OTUs. Clusters of unique OTUs are identified (dashed lines) and labelled according to cluster relationships in Figure 1 (red: no-agriculture; green: agriculture; purple: 19th century; fuchsia: modern). Taxonomic identifications for each OTU are listed in Table S8.

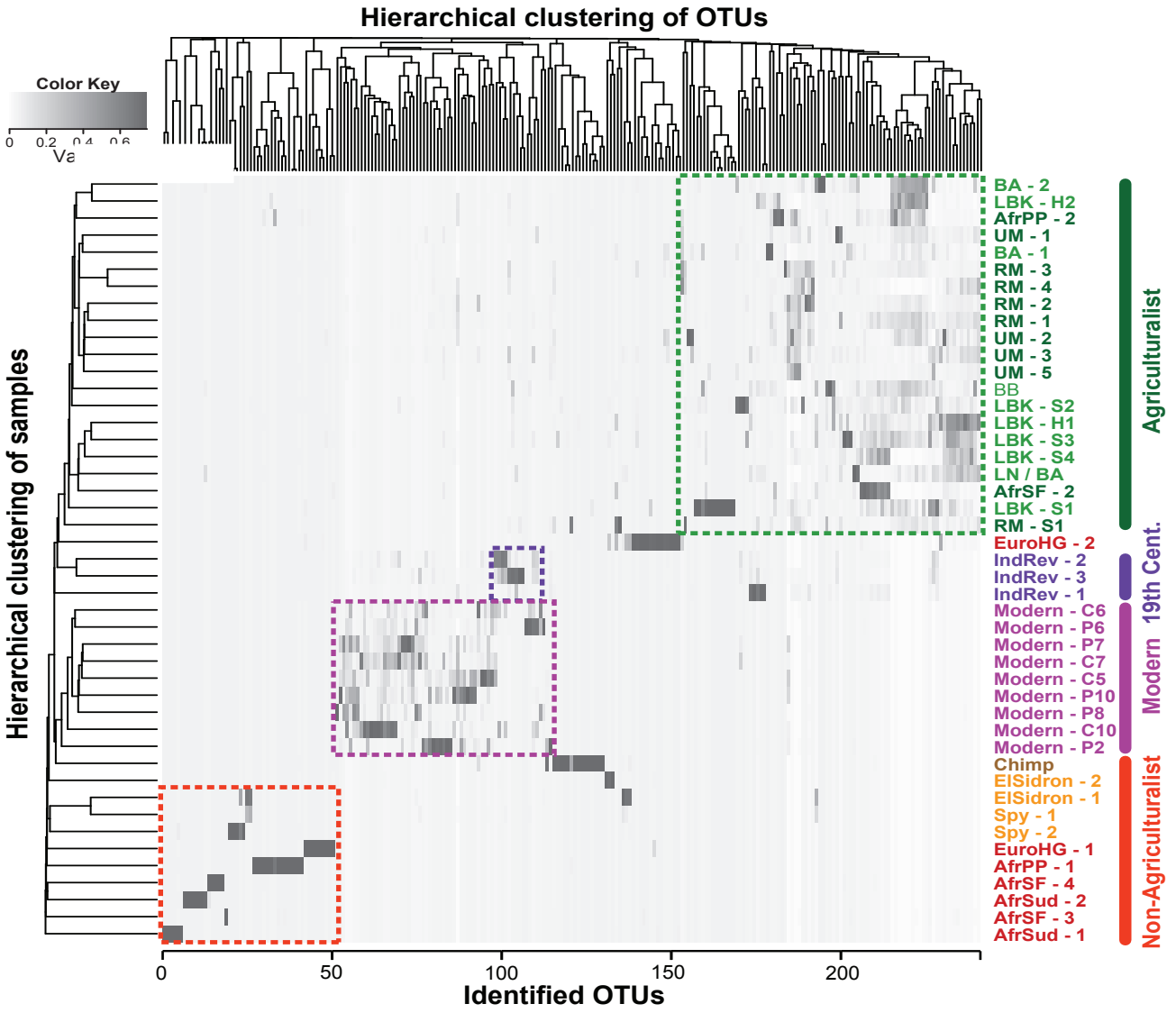


Table S3 – The taxa identified at a greater than 1% proportion in each sample were grouped according to bacterial phyla,. The proportion of organisms in each phyla was reported in this chart. The two phyla with the highest number of organisms was determined, and the cell containing that information is highlighted below.

Bacterial Phyla	No-Agriculture	Agriculture	19th Century	Modern
Acidobacteria	0.02	0.05	0.21	0.06
Actinobacteria	0.1	0.41	0	0
Bacteroidetes	0.04	0.01	0.43	0.36
Chlamydiae	0.02	0	0	0
Chloroflexi	0.02	0.05	0	0
Elusimicrobia	0.02	0	0	0
Euryarchaeota	0	0.15	0	0
Firmicutes	0.36	0.14	0.14	0.15
Fusobacterium	0	0	0	0.26
Nitrospirae	0	0.01	0	0
Planctomycetes	0	0.01	0	0
GAL15	0.02	0	0	0
Gemmatimonoadetes	0.02	0	0	0
Plactomycetes	0.04	0	0	0
Proteobacteria	0.28	0.09	0.21	0.13
Spiroachaetes	0	0	0	0.02
Synergistetes	0	0.04	0	0
Unidentified	0.06	0	0	0

Table S4 – Bacterial species identified from ancient dental calculus (Adler *et al.*, 2013) used to construct simulated metagenomes are listed.

Species genomes use to construct simulated metagenomes	Total number of sequences included	Percent of total metagenome
Agrobacterium tumefaciens 5a	2766360	3.82
Anaerolinea thermophila uni 1	1846052	2.55
Bacillus anthracis 52 g	2897088	4.00
Bacillus subtilis best7003	2323328	3.21
Bordetella pertussis 18323	1378624	1.90
Burkholderia cepacia gg4	2313732	3.19
Caldilinea aerophila dsm 14535 nbrc 104270	2441584	3.37
Campylobacter concisus 13826	1159324	1.60
Cardiobacterium hominis atcc 15826	1188668	1.64
Chlamydophila pneumoniae ar39	706596	0.98
Chlorobium limicola dsm 245	1520880	2.10
Enterococcus durans atcc 6056 ente dura atcc6056 v1	1709420	2.36
Escherichia coli e22	3044956	4.20
Fretibacterium fastidiosum	1021140	1.41
Fusobacterium necrophorum d12	1023220	1.41
Fusobacterium nucleatum 13 3c	1059856	1.46
Haemophilus aegyptius atcc 11116	1060140	1.46
Haemophilus haemolyticus hk386	1016300	1.40
Haemophilus influenzae 10810	1091756	1.51
Helicobacter pylori 2017	869468	1.20
Ignavibacterium album jcm 16511	1884708	2.60
Jonquetella anthropi e3 33 e1	831340	1.15
Klebsiella pneumoniae 120 1020	2673400	3.69
Lactobacillus acidophilus 30sc	1154960	1.59
Lactobacillus buchneri atcc 11577	1605508	2.22
Lactobacillus vaginalis atcc 49540	1031800	1.42
Leptotrichia buccalis c 1013 b	1109020	1.53
Listeria monocytogenes asm38292v1	1541648	2.13
Mycobacterium leprae tn	1604024	2.21
Mycobacterium tuberculosis 98 r604 inh rif em	1634900	2.26
Mycoplasma genitalium g37	281808	0.39
Neisseria bacilliformis atcc baa 1200	1103912	1.52
Neisseria gonorrhoeae 1291	1077240	1.49
Porphyromonas gingivalis atcc 33277	1331604	1.84
Prevotella denticola cris 18c a	1740700	2.40
Pseudomonas aeruginosa 18a	2203408	3.04
Pseudomonas fluorescens a506	2774324	3.83
Pyramidobacter piscolens w5455	1152084	1.59
Rhodobacter capsulatus b6	1384316	1.91
Staphylococcus aureus 04 02981	1412124	1.95
Staphylococcus caprae c87	1228496	1.70
Staphylococcus epidermidis 14 1 r1 se	1251648	1.73
Streptococcus mitis 11 5	1060000	1.46
Streptococcus mutans ua159	1100692	1.52
Streptococcus oralis atcc 35037 asm14856v1	1073252	1.48
Treponema denticola al 2	1539832	2.13
Treponema pallidum str fribourg blanc	622120	0.86
Yersinia pestis 113	2586256	3.57

Figure S8 – Simulated metagenomes (modern (circle) or ancient (square; damaged)) analyzed using four different software (DIAMOND (green); MALT (red); MetaPhylAn (blue); MG-RAST (orange)) were UPGMA clustered according to Bray Curtis distances calculated from genera within samples.

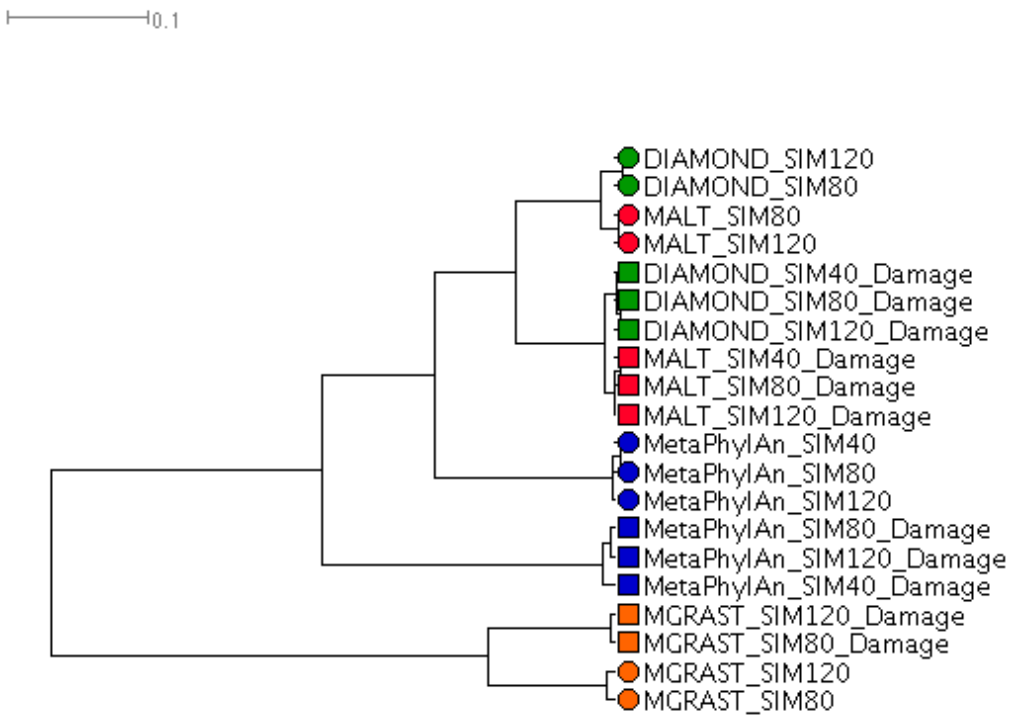


Figure S9 – Phyla identified in simulated metagenomes (modern or ancient) are shown for five different analysis programs: MALTX, DIAMOND, MetaPhylAn, and MG-RAST.

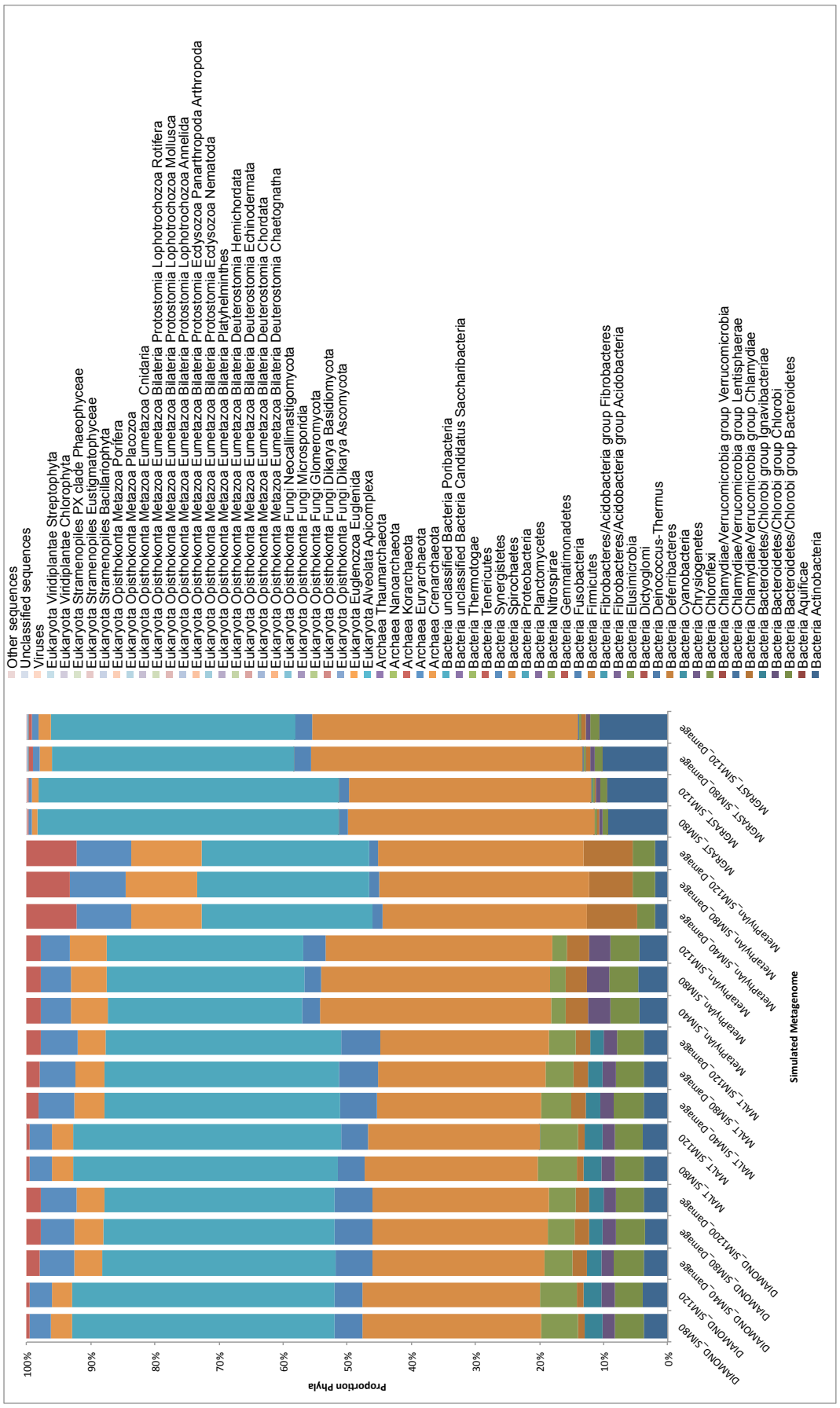


Figure S10 – MALTX analysis was completed on all available oral metagenomes in the MG-RAST repository, for both amplicon and shotgun sequencing methods. Modern calculus sequenced in this study (MODERN; teal) was compared to modern plaque samples (Ferre, *et al.*, 2012; blue; shotgun), a tracheal sample (Lazaridis, *et al.,* 2011; green; shotgun), and modern salivary samples (Cameron, *et al.*, 2015; red; amplicon) by UPGMA clustering samples based on Bray Curtis distances. Samples are labeled with the MG-RAST ascension number (mgm_XXXXXXX_MALT).

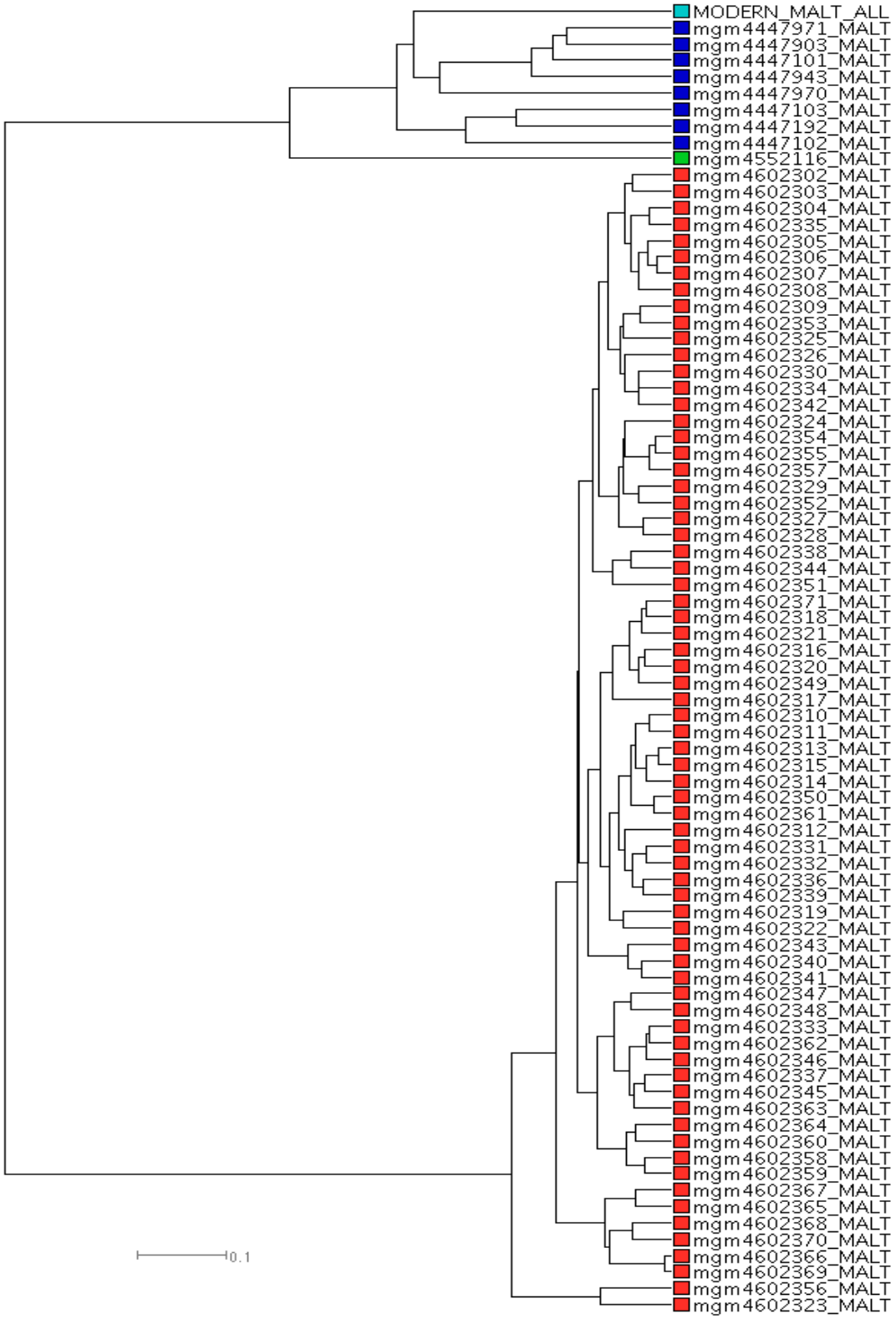


Figure S11 – Phyla identified by MALTX analysis in shotgun and amplicon oral datasets (Figure S15) obtained from MG-RAST are displayed in stacked bar plots.

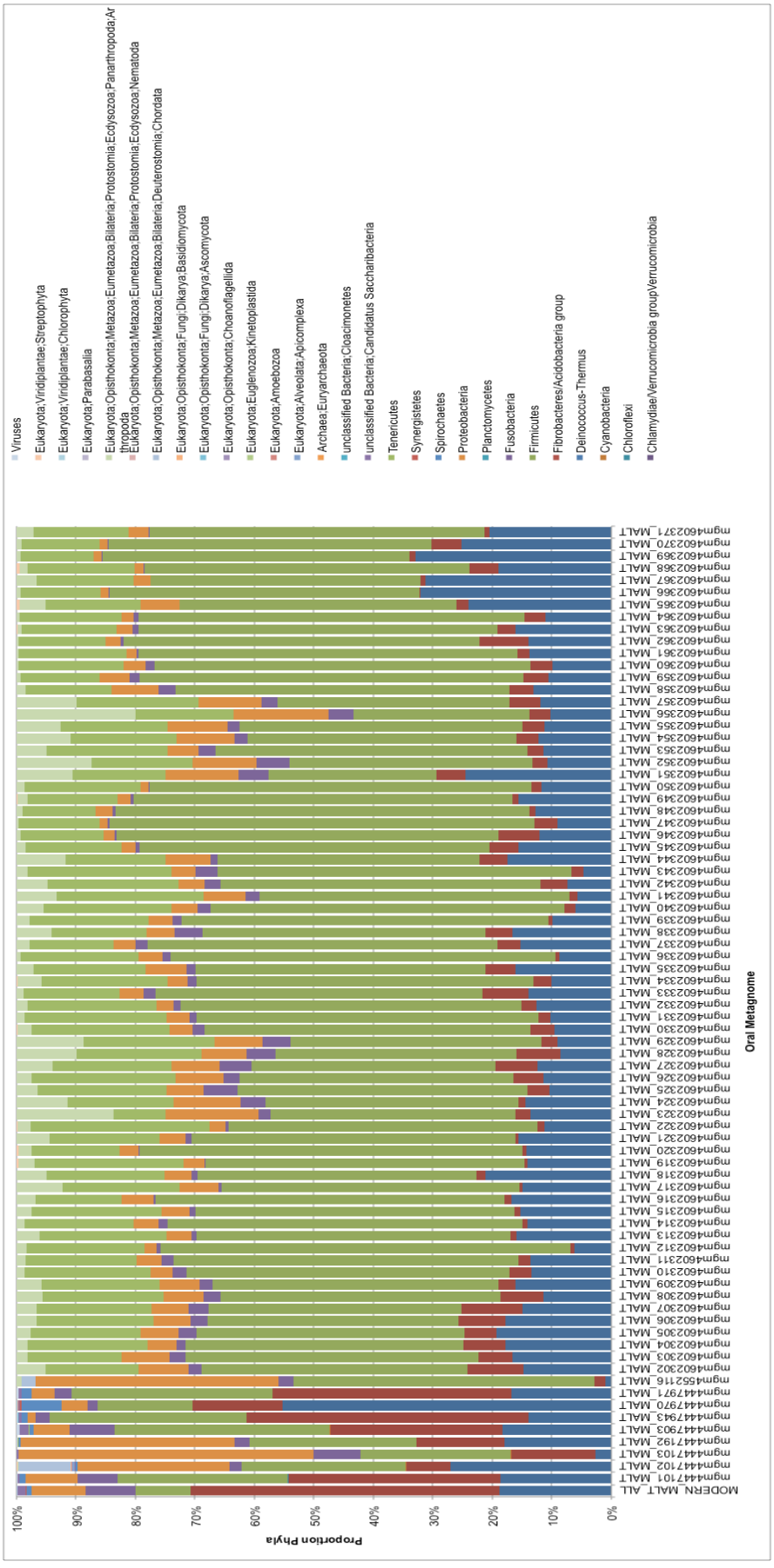


Table S5 – Shotgun metagenomic sequencing was performed on the following subset of samples, and the outputs were analysed using MALTX. The raw number of reads after demultiplexing is displayed, as well as the number of reads present after adapter trimming and filtering species that were identified within the EBCs.

Samples	Raw Data # Reads	Post-Filtering # Reads	% Removed
Chimp	18680535	17214005	8%
ElSidron1	56584638	49943977	12%
ElSidron2	66905980	47820005	29%
Spy I	69901550	3996189	94%
Spy II	87504507	17388077	80%
WarB61	94679298	945484	99%
WarG12	77177157	767283	99%
Modern	44000223	26735211	39%
EBC1	6350329	53309	99%
EBC2	663351	9070	99%
EBC3	238422	238395	0%
BraccWater	14533539	2383467	84%
GrndWater	76271556	8191610	89%
GrassSoil	93579626	35243604	62%
ForestSoil	50544126	6187071	88%
Total Reads	757614837	217116757	71%

Figure S12 – Unfiltered prokaryotic phyla identified from 16S rRNA results and shotgun sequencing results (MALTX) are compared.

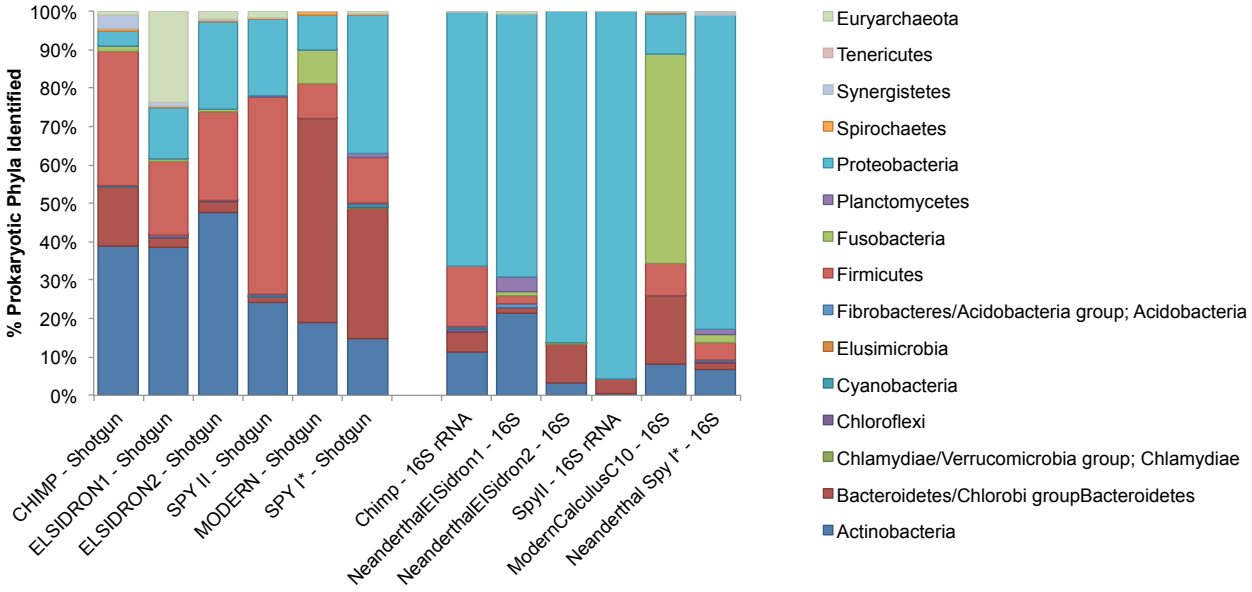


Figure S13 – Archaeal and bacterial composition is plotted for filtered and unfiltered 16S rRNA and shotgun datasets, as well as 16S rRNA reads identified by GraftM.

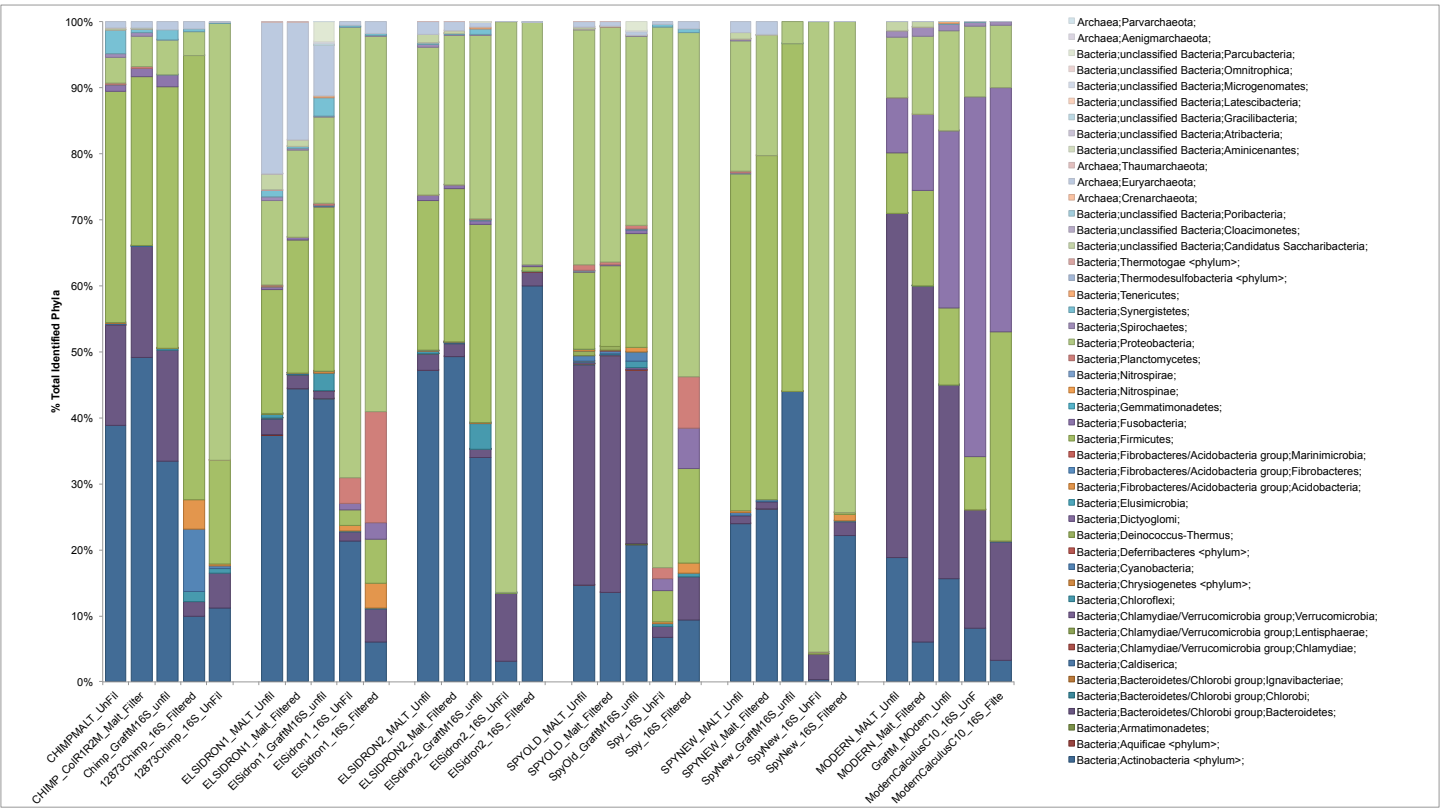


Table S6 – 16S rRNA reads identified within deeply-sequenced shotgun datasets were identified by GraftM, and the average read length was calculated using the read_length.py script in bbmap.

	Total Reads	Maximum Length	Minimum Length	Average Length	Median Length	Mode Length
Chimpanzee	8639	166	44	75.4	70	60
El Sidron 1	20035	180	44	70.1	60	60
El Sidron 2	33042	181	44	74.8	70	60
Spy 1	7307	179	44	79.7	70	70
Spy 2	2256	141	49	89.6	70	60
Warinner B61	327	604	400	138.2	120	100
Warinner G12	232	431	100	137.3	120	100
Modern	33612	186	43	89.8	80	70

Figure S14 – Raw shotgun sequences were analyzed by MALTX and by MG-RAST. The resulting, unfiltered bacterial sequences from each analysis are shown below.

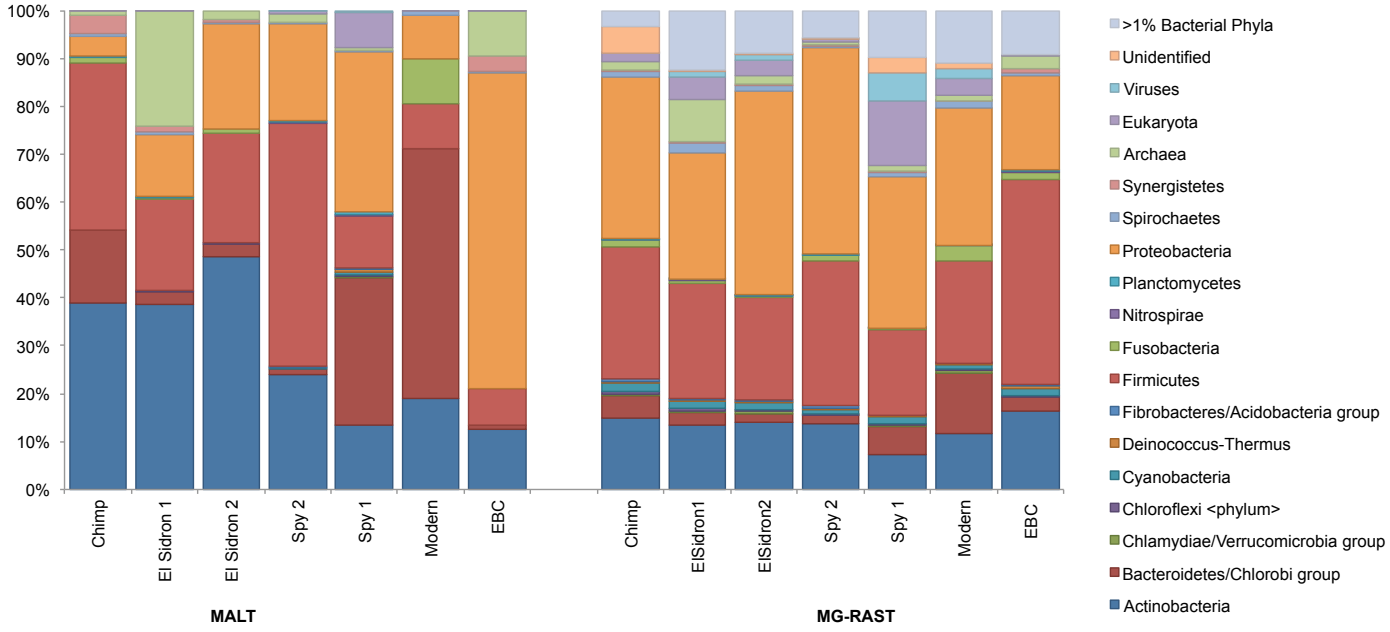


Figure S15 – MALTX was utilized to identify all reads in ancient, historic, and modern metagenomic samples, which were compared to extraction blank controls (EBCs) and environmental samples.

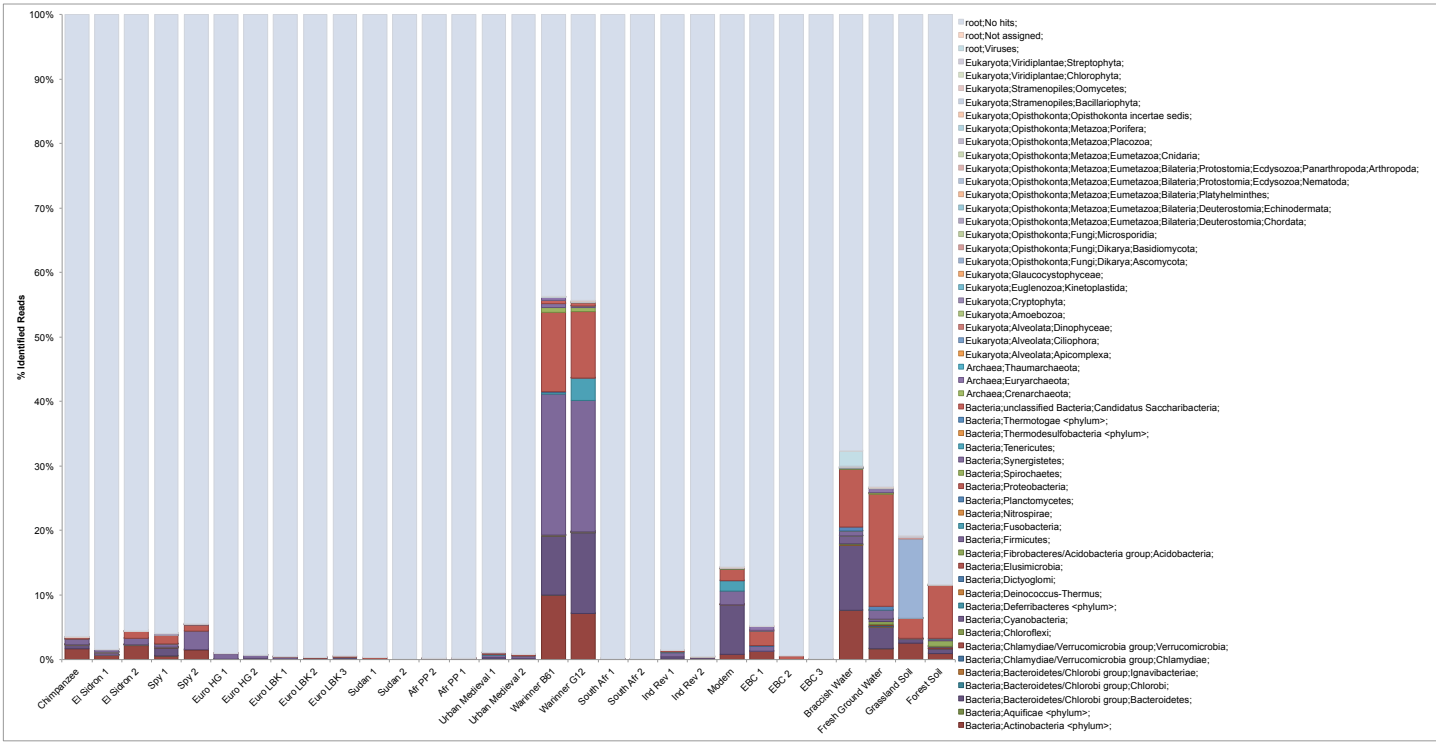


Figure S16 – Identified reads using MALTX, classified at the phyla level, for both dental calculus samples, extraction blank controls (EBCs), and environmental samples. Ancient dental calculus samples are graphs in order of age, with the oldest specimens listed on the left.

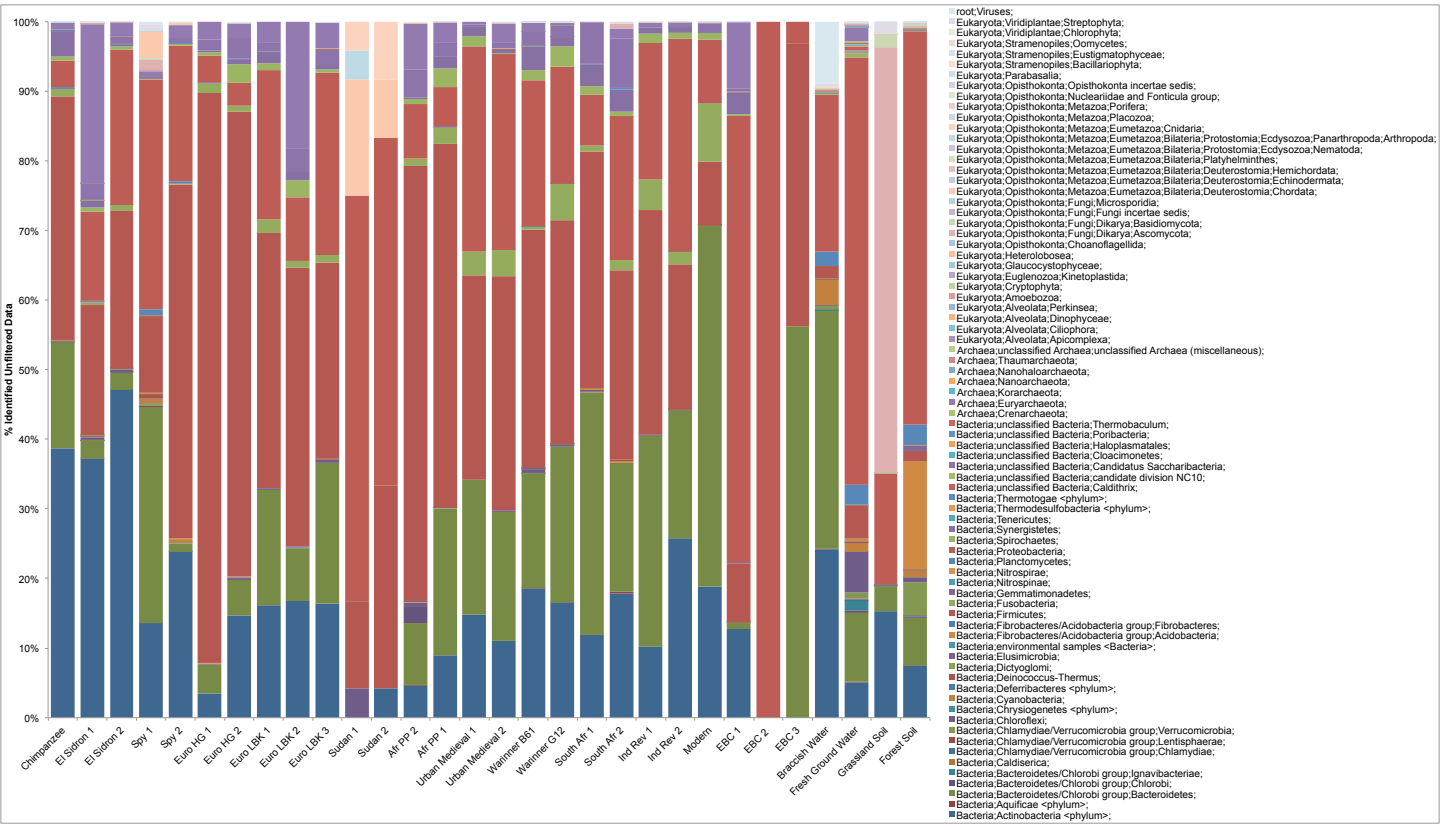
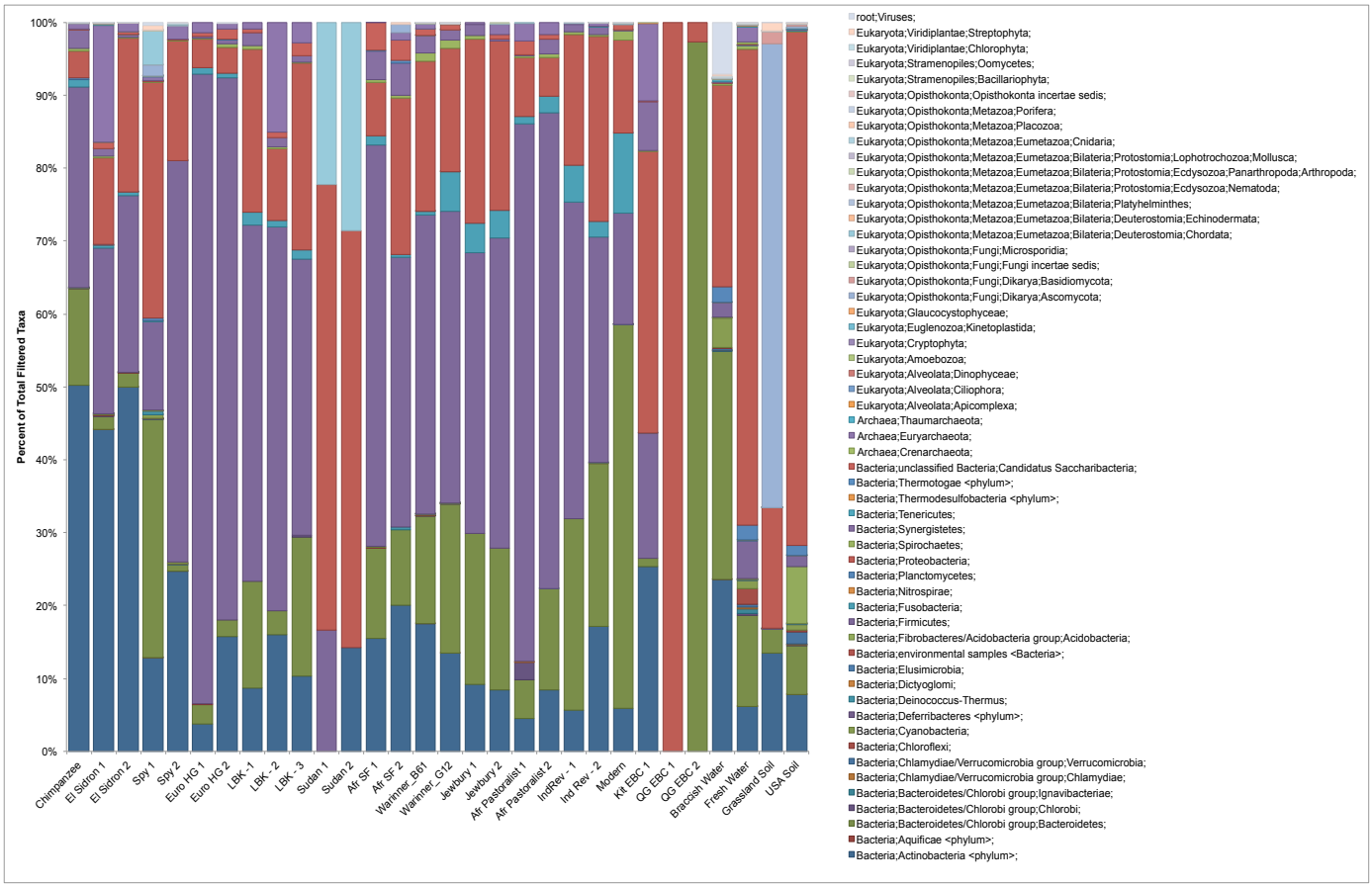


Figure S17 – (A) Identified reads using MALTX were filtered to remove reads corresponding to species identified in extraction blank controls from QG DNA extractions and environmental controls. (B) Filtered data was summarized to examine only phyla of archaea and bacteria typically found in the modern oral cavity. Dental calculus samples are graphed in order of age.

A



B

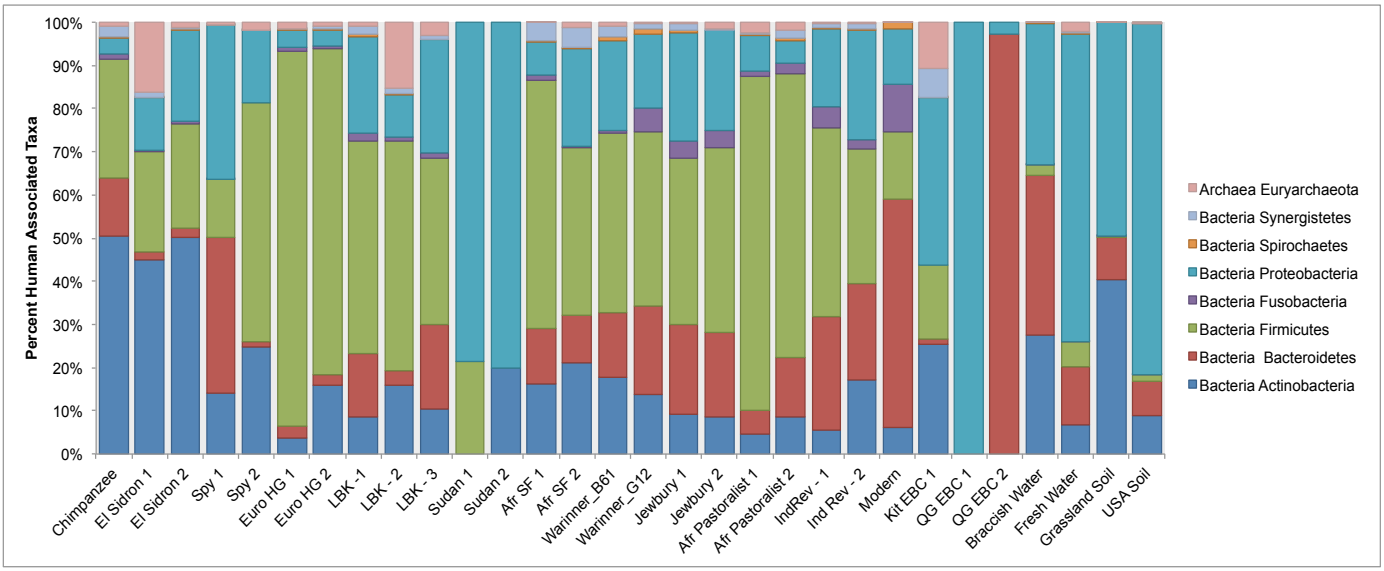


Table S7 – The total number of reads analyzed by MALTX (nucleotide alignment) in each metagenomic sample was calculate prior to and after filtering species identified in QG extraction blank control reads and environmental samples (BraccWater, GrndWater, GrassSoil, ForestSoil) from ancient dental calculus samples.

Samples:	Raw Data # Reads	Post-Filtering # Reads	% Removed
Chimp	18680535	17214005	8%
ElSidron1	56584638	49943977	12%
ElSidron2	66905980	47820005	29%
Spy1	87504507	17388077	80%
Spy2	69901550	3996189	94%
EuroHG1	145907	145670	0%
EuroHG2	93208	93046	0%
EuroLBK1	132671	39379	70%
EuroLBK2	258241	125359	51%
EuroLBK3	346073	158045	54%
Sudan1	203099	5240	97%
Sudan2	284254	81138	71%
SouthAfr1	3649738	3390649	7%
SouthAfr2	654275	308064	53%
WarB61	94679298	945484	99%
WarG12	77177157	767283	99%
UrbMed1	197853	51720	74%
UrbWar2	187667	70906	62%
AfrPP2	1105368	820741	26%
AfrPP1	6292653	2641418	58%
IndRev1	116542	115772	1%
IndRev2	11702696	11435338	2%
Modern	44000223	26735211	39%
EBC1	6350329	53309	99%
EBC2	663351	9070	99%
EBC3	238422	238395	0%
BraccWater	14533539	2383467	84%
GrndWater	76271556	8191610	89%
GrassSoil	93579626	35243604	62%
ForestSoil	50544126	6187071	88%
Total Reads	309905088	236512864	24%

Figure S18 – Identified reads by DIAMOND analysis are shown for the 60 phyla identified in ancient and modern dental calculus, extraction blank controls, and environmental samples.

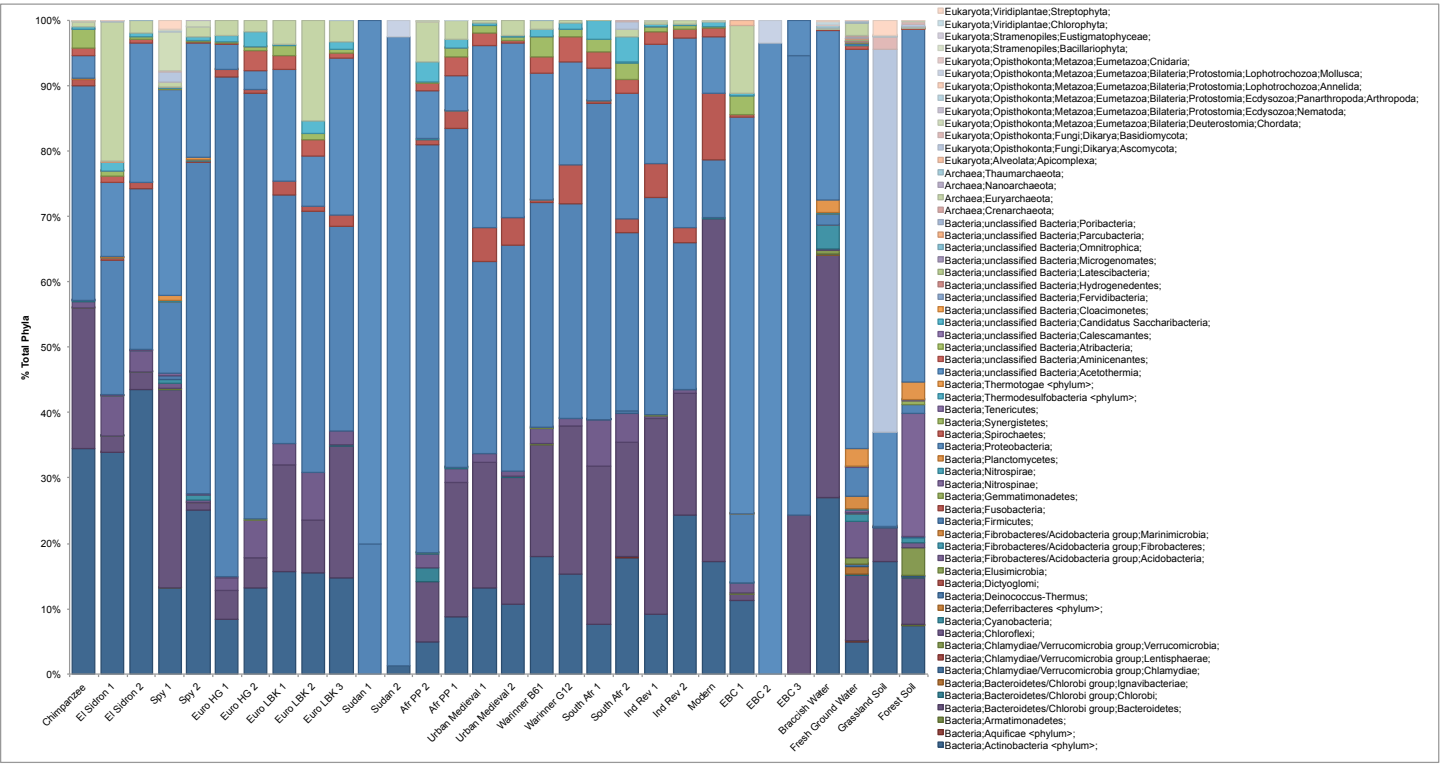


Figure S19 – Reads identified by DIAMOND analysis were filtered to remove species identified in extraction blank controls (EBCs) and environmental controls from ancient, historic, and modern calculus samples.

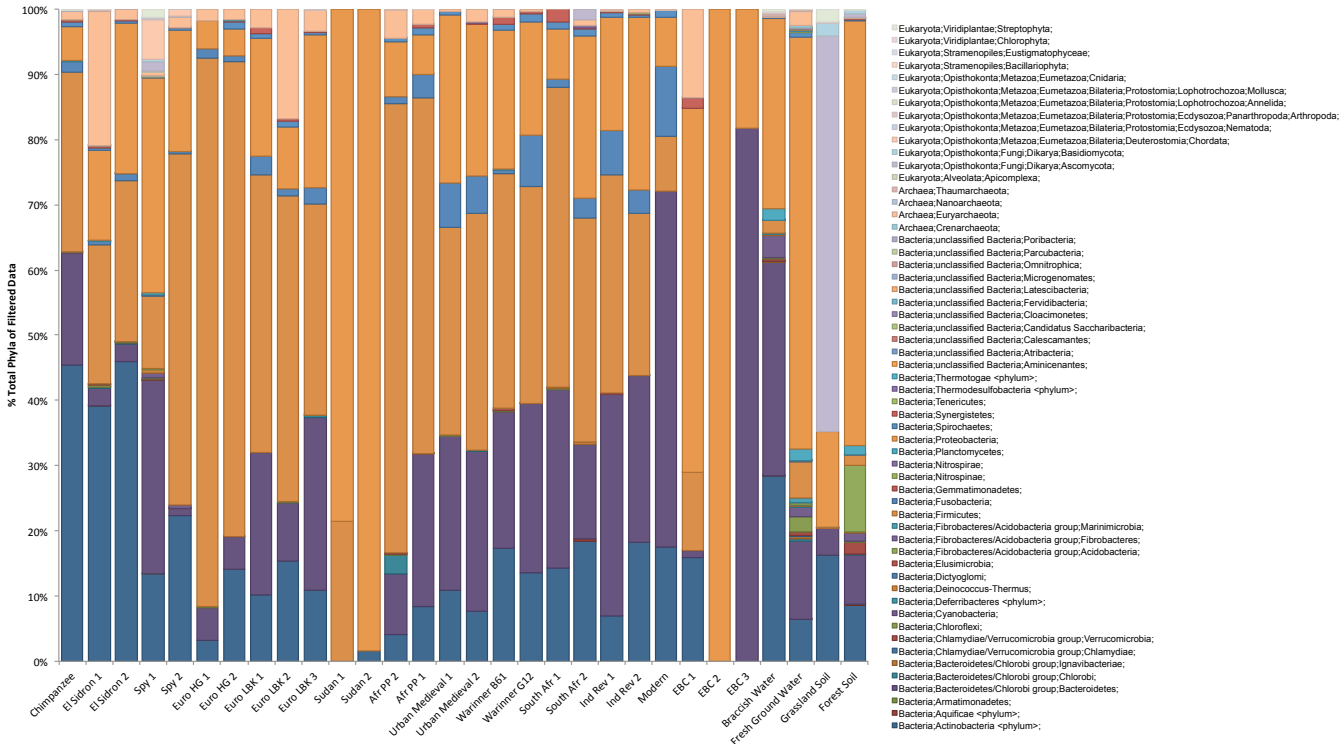


Figure S20 – Alpha diversity from deeply sequenced unfiltered shotgun datasets was calculated using Shannon-Weaver (A) and Simpson's Reciprocal (B) indexes.

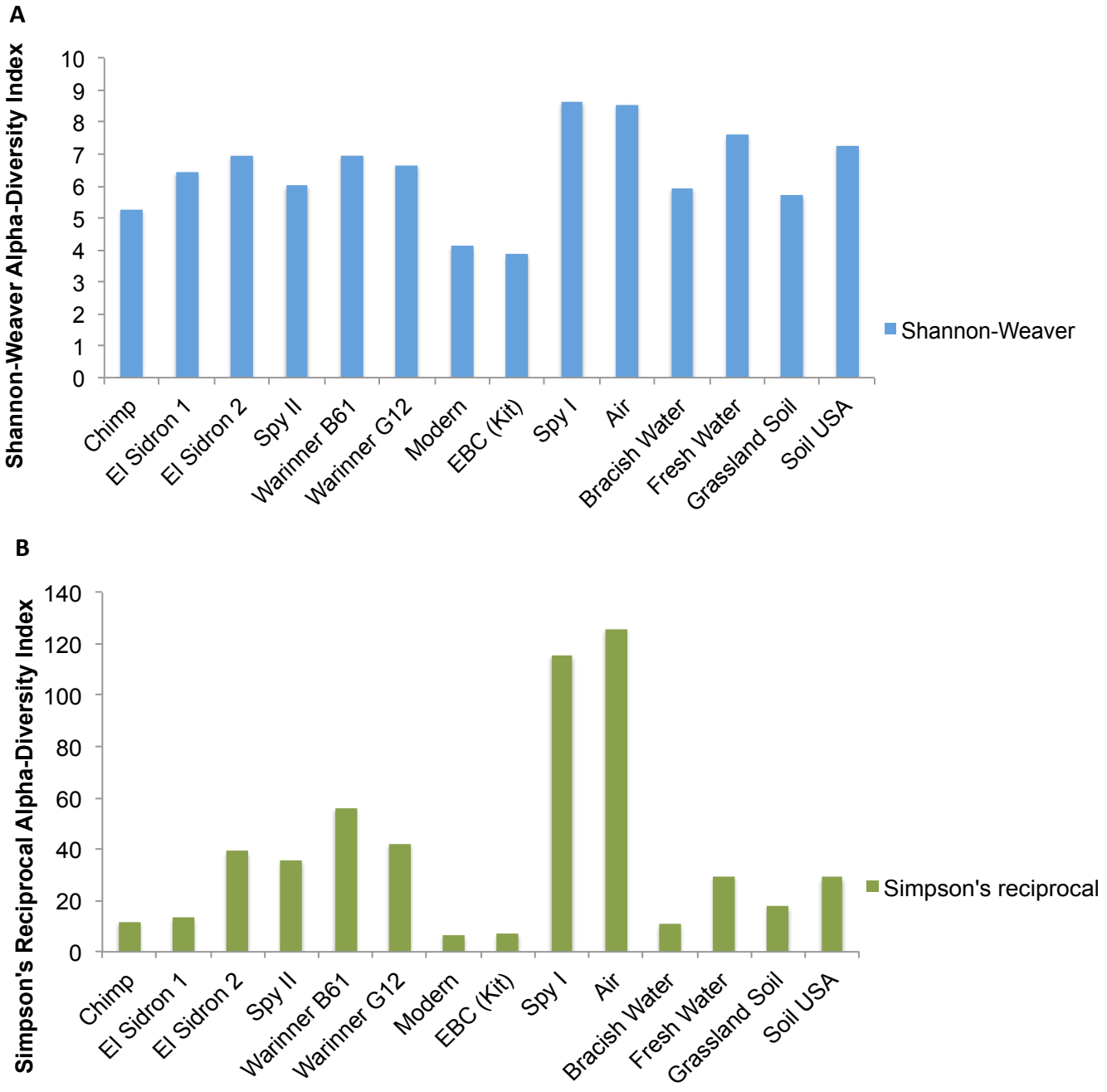
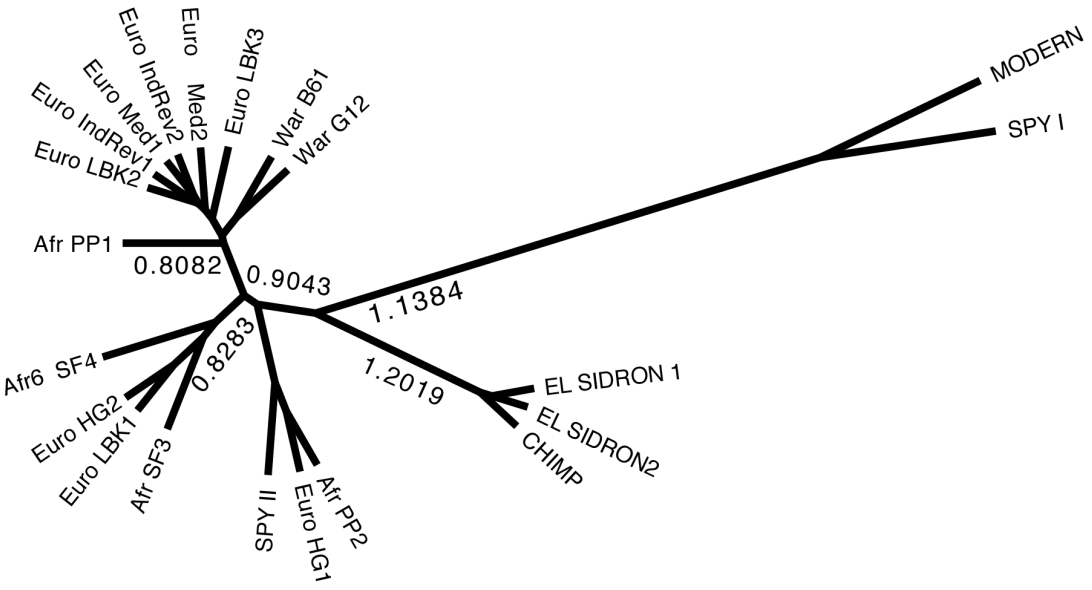


Figure S21 – UPGMA clustering of Bray-Curtis values calculated from filtered rarefied shotgun data. (**B**) The groups were largely split based on their differences in the proportion of Gram-positive and Gram-negative phyla in shotgun datasets was plotted for each group (chimpanzee and modern human, n=1; Neandertals, n=3). Error bars represent standard deviation.

A



B

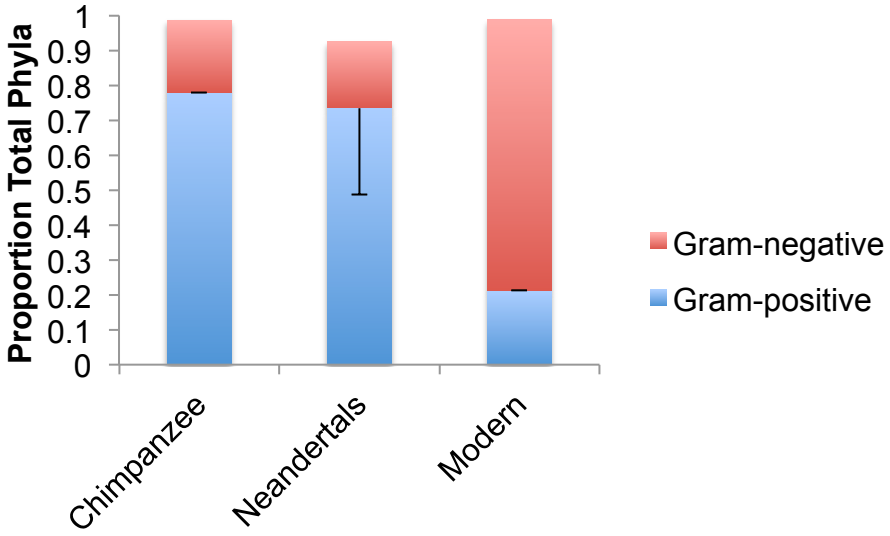


Table S8 – Species shared with Neandertals, chimpanzee and modern human samples were identified in MEGAN5 from the MALTX analysis. Percentage of total identified species were calculated and displayed for each sample.

	Chimpanzee	El Sidron 1	El Sidron 2	Spy 1	Spy 2	Modern
Actinobacteria - *Actinomyces naeslundii*	0.22%	2.64%	8.36%	0.34%	0.01%	0.29%
Actinobacteria - *Actinomyces odontolyticus*	0.15%	0.63%	0.91%	0.24%	1.02%	0.11%
Actinobacteria - *Actinomyces sp. ICM47*	0.09%	0.30%	0.37%	0.15%	0.05%	0.01%
Actinobacteria - *Actinomyces sp. oral taxon 178*	0.67%	1.40%	0.57%	0.12%	0.01%	0.66%
Actinobacteria - *Actinomyces sp. oral taxon 849*	1.05%	4.32%	11.47%	0.45%	1.39%	0.26%
Actinobacteria - *Corynebacterium durum*	0.01%	0.06%	0.04%	0.03%	1.27%	0.05%
Bacteroidetes - *Prevotella intermedia*	0.42%	0.38%	0.08%	0.01%	1.23%	0.06%
Firmicutes - *Gemella haemolysans*	0.02%	0.03%	0.25%	0.06%	0.65%	0.07%
Firmicutes - *Streptococcus sanguinis*	0.10%	1.00%	3.78%	1.40%	9.46%	0.08%
Firmicutes - *Mogibacterium sp. CM50*	13.20%	2.74%	3.38%	0.97%	13.88%	0.03%
Firmicutes - *Dorea longicatena*	0.01%	0.04%	0.04%	0.03%	0.12%	0.01%
Firmicutes - *Johnsonella ignava*	0.21%	0.67%	2.69%	0.35%	0.02%	0.47%
Firmicutes - *Lachnoanaerobaculum saburreum*	0.38%	0.19%	0.37%	0.07%	0.06%	0.23%
Firmicutes - *Lachnospiraceae oral taxon 107*	0.15%	0.15%	0.25%	0.03%	0.60%	0.17%
Firmicutes - *Filifactor alocis*	0.33%	0.37%	0.28%	0.07%	0.97%	0.02%
Firmicutes - P*eptostreptococcus anaerobius*	0.02%	0.12%	0.04%	0.03%	1.21%	0.03%
Firmicutes - *Peptostreptococcus stomatis*	0.11%	3.86%	0.77%	0.44%	10.25%	0.52%
Spirochaetes - *Treponema vincentii*	0.27%	0.32%	0.15%	0.04%	0.41%	5.98%

Figure S22 – MapDamage analysis was performed on the reads mapping to shared oral bacterial species in Neandertals and a modern human. The percent of C-T mutations (A) or read length (B) calculated from mapped reads of each sample is graphed for ten conserved species.

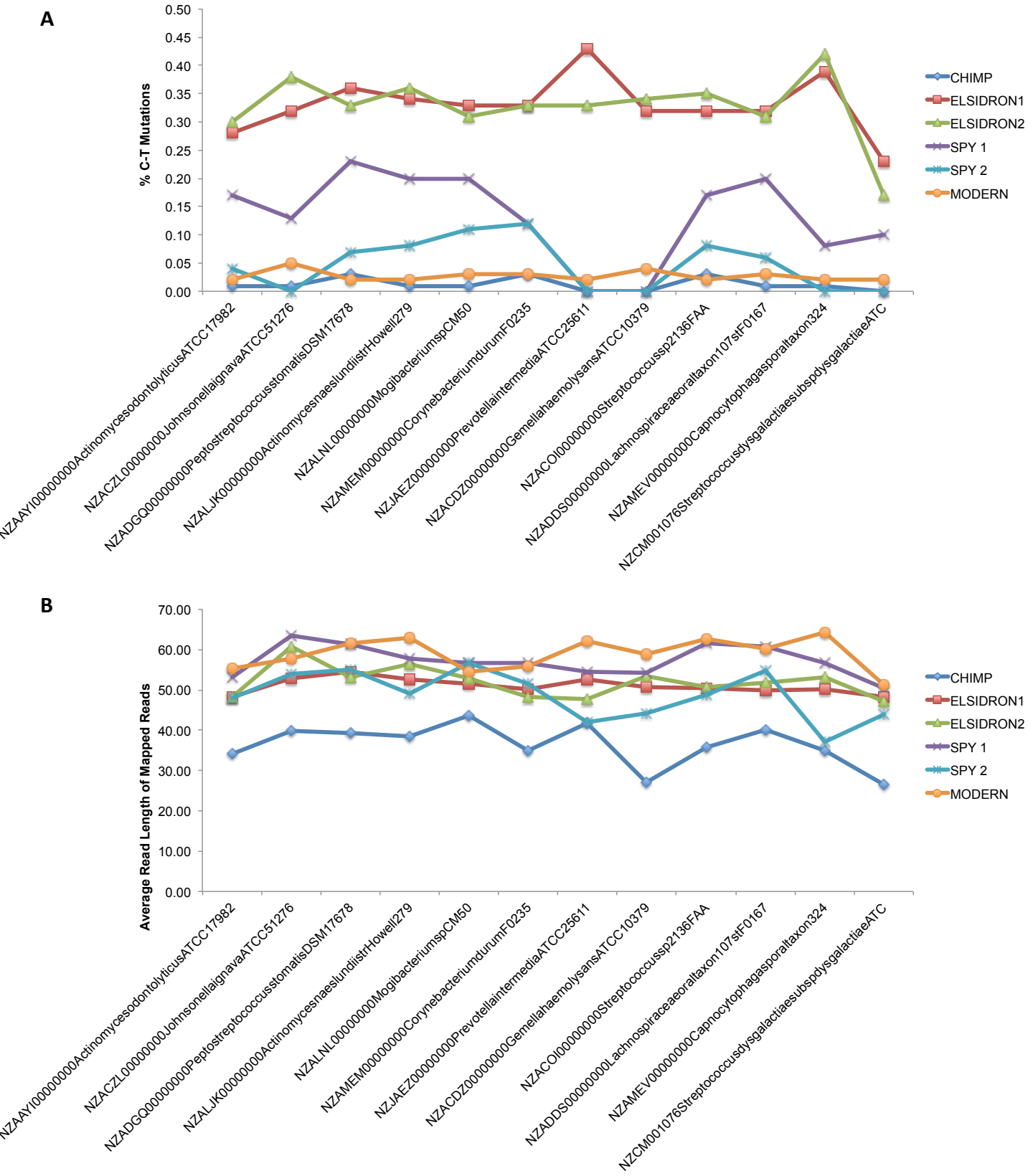


Table S9 – Identification of Neandertal bacterial pathogens. Species identifications from MALTX analysis were screened for known human bacterial pathogens. The number of confirmed hits to each species is given for all deeply sequenced metagenomes (chimpanzee, Neandertals, and modern human), other published ancient oral metagenomes (WarB61 and G12), extraction blank control (EBC) samples from DNA extraction kits and homemade DNA QG extraction methods, and environmental samples (soil and water). * indicates that the bacterial species is closely related to several common oral species.

Pathogens	**Dental Calculus**								**Extraction Blank Controls**			**Environmental Samples**			
Pathogens in Ancient Dental Calculus	**Chimp**	**El Sidron 1**	**El Sidron 2**	**Spy 1**	**Spy 2**	**War B61**	**War G12**	**Modern**	**Kit EBC**	**QG EBC 1**	**QG EBC 2**	**Bracc. Water**	**Grnd Water**	**GrssInd Soil**	**Forest Soil**
Bordetella parapertussis	0	23	60	24	0	13	0	0	0	0	0	0	0	0	0
Pasteurella multocida	0	20	0	13	0	41	36	0	0	0	0	0	0	0	0
*Neisseria gonorrhoeae**	0	0	116	41	90	58	180	162	0	0	0	0	0	0	0
*Streptococcus pyogenes**	0	37	273	14	0	0	13	0	0	0	0	0	0	0	0
*Corynebacterium diphtheriae**	82	54	90	60	0	70	68	221	0	0	0	0	0	0	0
Pathogens in DNA Extraction Kit Reagents															
Treponema denticola	1506	2826	4502	81	0	2984	10454	6618	8	0	0	24	222	0	21
Porphyromonas gingivalis	15982	785	652	70	0	379	2230	1263	5	0	0	23	277	0	0
Tannerella forsythia	25634	3113	7359	67	342	9286	13605	27860	22	0	0	46	348	0	29
Salmonella enterica	0	29	38	214	0	26	14	0	69	0	0	75	191	131	50
Clostridium botulinum	107	184	236	89	0	83	70	170	1	0	0	52	557	0	40
Pathogens in Environmental Samples															
Streptococcus mutans	18	194	550	95	23	158	133	126	0	0	0	0	55	1388	0
Staphylococcus aureus	0	59	207	102	0	40	28	80	0	0	0	132	186	0	31
Clostridium tetani	0	48	44	0	0	19	17	0	0	0	0	0	84	0	0
Neisseria meningitidis	0	84	472	76	0	175	302	2073	0	0	0	18	84	0	22
Helicobacter pylori	0	70	91	434	483	24	33	93	0	0	0	52	162	0	47
Edwardsiella tarda	0	0	58	12	0	0	0	0	0	0	0	19	0	0	0
Klebsiella pneumoniae	0	14	0	421	0	12	11	204	0	0	0	27	235	0	28
Serratia marcescens	0	0	104	10	0	14	15	0	0	0	0	14	74	0	28
Yersinia pestis	0	0	0	18	0	0	0	0	0	0	0	15	0	0	0
Legionella pneumophila	0	34	112	112	0	0	0	0	0	0	0	273	879	0	220
Pseudomonas aeruginosa	0	28	52	164	0	22	14	75	0	0	0	93	4185	132	81
Vibrio cholerae	0	67	99	92	0	23	17	102	0	0	0	215	603	128	81
Listeria monocytogenes	29	25	80	15	0	19	13	0	0	0	0	0	86	0	0
Campylobacter jejuni	72	57	61	0	0	8	23	116	0	0	0	43	141	0	16
Escherichia coli	94	156	379	647	3	62	37	180	0	0	0	149	562	139	116
Mycobacterium tuberculosis	166	97	159	109	0	13	13	152	0	0	0	547	779	721	646
Fusobacterium nucleatum	3941	675	2022	36	0	116	3950	79947	0	0	0	27	105	0	0

Table S10 – Identification of common oral pathogens in dental calculus, as well as extraction blank control (EBC) samples and environmental controls. The number of reads identified by MALTX and the percentage of the total reads identified are given, alongside the factor increase of these reads in Neandertals over either the deeply sequenced kit EBC (EBC1) or fresh ground water.

	Porphyromonas gingivalis				*Tannerella forsythia*				*Treponema denticola*				*Streptococcus mutans*			
	Reads Identified	% Identified	Fct Inc over GF	Fct Inc over EBC	Reads Identified	% Identified	Fct Inc over Ground Water	Fct Inc over EBC	Reads Identified	% Identified	Fct Inc over Ground Water	Fct Inc over EBC	Reads Identified	% Identified	Fct Inc over Ground Water	Fct Inc over EBC
Chimpanzee	15982	0.85555%	57.69663	3196.403428	25634	0.13722%	73.66099	1165.178319	1506	0.00806%	6.78380	188.250114	**18**	**9.63571E-07**	**0.327272866**	na
El Sidron 1	785	0.00420%	2.83393	157.0001684	3113	0.01666%	8.94541	141.4995751	2826	0.01513%	12.72976	353.2502139	194	1.03851E-05	3.527274115	na
El Sidron 2	652	0.00349%	2.35379	130.4001399	7359	0.03939%	21.14657	334.4989954	4502	0.02410%	20.27933	562.7503408	550	2.94424E-05	10.00000393	na
Spy 1	70	0.00037%	0.25271	14.00001502	67	0.00036%	0.19253	3.045445399	81	0.00043%	0.36487	10.12500613	95	5.08551E-06	1.727273407	na
Spy 2	0	0.00000%	0.00000	0	342	0.00183%	0.98276	15.54540786	0	0.00000%	0.00000	0	23	1.23123E-06	0.418181983	na
Warinner B61	379	0.00203%	1.36823	75.80000813	9286	0.04971%	26.68393	422.0896415	2984	0.01597%	13.44147	373.0002259	158	8.458E-06	2.872728403	na
Warinner G12	2230	0.01194%	8.05052	446.0004784	13605	0.07283%	39.09486	618.4072337	10454	0.05596%	47.09020	1306.750791	133	7.11971E-06	2.418182769	na
Modern	1263	0.00676%	4.55956	252.6002709	27860	0.14914%	80.05754	1266.359833	6618	0.03543%	29.81088	827.2505501	126	6.74499E-06	2.290909992	na
EBC1 (Kit)	5	0.00003%	0.01805	1.000001073	22	0.00012%	0.06322	0.999996997	8	0.00004%	0.03604	1.000000606	0	0	0	na
EBC 2 (QG)	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	0	0	0	na
EBC 3 (QG)	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	0	0	0	na
Braccish Water	23	0.00012%	0.08303	4.600004934	46	0.00025%	0.13218	2.090902812	24	0.00013%	0.10811	3.000001817	0	0	0	na
Ground Water	277	0.00148%	1.00000	55.40005942	348	0.00186%	1.00000	15.81813431	222	0.00119%	1.00000	27.7500168	55	2.94424E-06	1.000000393	na
Grassland Soil	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	0	0.00000%	0.00000	0	1388	7.43019E-05	25.23637356	na
Forest Soil	0	0.00000%	0.00000	0	29	0.00016%	0.08333	1.318177859	21	0.00011%	0.09459	2.62500159	0	0	0	na

Table S11 – MapDamage analysis was performed on the reads mapping to oral bacterial pathogen genomes in chimpanzee, Neandertal, and modern human specimens.

Ancient Library	Reference	All Mapped Reads	Mapped Reads >Q30	Deduplicted Mapped Reads	5' C-T Proportion	3' G-A Proportion	Increase Purines Before Start	Decrease Pyrimindines Before Start	Average Length	St Dev Length	Cytosine Deamination in Dbl Strds (DeltaD)	Cytosine Deamination in Single Strds (DeltaS)	Avg Length of Single Strd Overhangs (Lambda)	Ratio of DeltaD to DeltaS
DCCHIMP	NC002967TreponemadenticolaATCC35405	9353	2640	177	0	0.04	0.62	3.04	32.66	13.67	0.03	0.29	0.7	0.103448
DCELSIDRON1	NC002967TreponemadenticolaATCC35405	30753	18556	11596	0.38	0.44	1.57	0.53	52.66	13.12	0.02	1	0.34	0.02
DCELSIDRON2	NC002967TreponemadenticolaATCC35405	32689	20390	1978	0.33	0.39	1.63	0.53	57.16	21.01	0.03	0.98	0.39	0.030612
DCSPYNEW	NC002967TreponemadenticolaATCC35405	437	2	2	0	0	0.5		36.5	9.19	0.1	0.47	0.56	0.212766
DCSPYOLD	NC002967TreponemadenticolaATCC35405	1341	111	75	0	0.08	0.96	1.07	36.96	21.1	0.04	0.32	0.65	0.125
DCMODERN	NC002967TreponemadenticolaATCC35405	19351	7942	7572	0.03	0.02	0.89	1.13	59.94	25.57	0.03	0.15	0.71	0.2
DCCHIMP	NC004350StreptococcusmutansUA159	9554	412	62	0	0	0.59	3.75	26.84	5.1	0.09	0.39	0.51	0.230769
DCELSIDRON1	NC004350StreptococcusmutansUA159	17931	1026	617	0.25	0.26	1.49	0.6	46.28	12.2	0.01	0.9	0.45	0.011111
DCELSIDRON2	NC004350StreptococcusmutansUA159	31849	2676	259	0.26	0.25	1.32	0.68	46	15.01	0	0.83	0.5	0
DCSPYNEW	NC004350StreptococcusmutansUA159	4420	440	9	0	0	5	0.5	39.89	9.03	0.05	0.41	0.59	0.121951
DCSPYOLD	NC004350StreptococcusmutansUA159	3778	368	128	0.11	0	1.42	0.55	49.38	19.86	0.01	0.34	0.74	0.029412
DCMODERN	NC004350StreptococcusmutansUA159	19696	1472	1391	0.02	0.03	0.89	1.12	46.45	19.62	0.01	0.23	0.79	0.043478
DCCHIMP	NC010729PorphyromonasgingivalisATCC33277	126008	100828	2350	0.02	0.01	0.87	1.21	46.17	19.56	0.03	0.09	0.56	0.333333
DCELSIDRON1	NC010729PorphyromonasgingivalisATCC33277	22238	12985	7493	0.44	0.51	1.61	0.52	50.04	12.51	0	1	0.33	0
DCELSIDRON2	NC010729PorphyromonasgingivalisATCC33277	18765	7547	720	0.42	0.35	1.62	0.57	48.19	16.68	0.01	0.97	0.36	0.010309
DCSPYNEW	NC010729PorphyromonasgingivalisATCC33277	175	7	6	0	0	3	0.6	38.33	21.13	0.04	0.44	0.59	0.090909
DCSPYOLD	NC010729PorphyromonasgingivalisATCC33277	2348	145	63	0	0.08	0.84	1.23	51.17	29.4	0.01	0.37	0.75	0.027027
DCMODERN	NC010729PorphyromonasgingivalisATCC33277	42018	13938	13153	0.02	0.02	0.84	1.17	60.93	26.04	0.01	0.12	0.69	0.083333
DCCHIMP	NC016610Tannerellaforsythia92A2	220888	166503	4325	0.02	0.01	0.81	1.35	43.68	17.93	0.03	0.01	0.47	3
DCELSIDRON1	NC016610Tannerellaforsythia92A2	71435	50388	30085	0.42	0.46	1.63	0.5	49.21	11.84	0.01	1	0.34	0.01
DCELSIDRON2	NC016610Tannerellaforsythia92A2	94682	66842	6209	0.41	0.42	1.62	0.53	49.23	15.8	0.01	1	0.34	0.01
DCSPYNEW	NC016610Tannerellaforsythia92A2	181	8	8	0.33	0	1.2	0.67	41.25	16.14	0.04	0.58	0.49	0.068966
DCSPYOLD	NC016610Tannerellaforsythia92A2	2671	347	152	0.09	0.12	1	1	45.25	19.59	0.01	0.44	0.54	0.022727
DCMODERN	NC016610Tannerellaforsythia92A2	131223	98558	93666	0.01	0.01	0.84	1.17	64.97	27.22	0.01	0.07	0.57	0.142857

Table S12 – MapDamage analysis was performed on the reads mapping to pathogenic bacterial genomes in chimpanzee, Neandertal, and modern human specimens.

Ancient Library	Reference	All Mapped Reads	Mapped Reads >Q30	Deduplicted Mapped Reads	5' C-T Proportion	3' G-A Proportion	Increase Purines Before Start	Decrease Pyrimindines Before Start	Average Length	St Dev Length	Cytosine Deamination in Dbl Strds (DeltaD)	Cytosine Deamination in Single Strds (DeltaS)	Avg Length of Single Strd Overhangs (Lambda)	Ratio of DeltaD to DeltaS
DCCHIMP	NC009567Haemophilusinfluenzae PittGG	6645	387	43	0	0	0.64	2.86	26.67	3.85	0.06	0.32	0.64	0.1875
DCELSIDRON1	NC009567Haemophilusinfluenzae PittGG	11510	477	319	0.18	0.46	1.68	0.49	44.49	11.21	0.01	0.9	0.45	0.011111
DCELSIDRON2	NC009567Haemophilusinfluenzae PittGG	18201	997	127	0.47	0.32	1.5	0.61	41.23	12.63	0.02	0.84	0.46	0.02381
DCSPYNEW	NC009567Haemophilusinfluenzae PittGG	1591	523	5	0	0	0.75	2	39.2	13.72	0.03	0.42	0.65	0.071429
DCSPYOLD	NC009567Haemophilusinfluenzae PittGG	3124	220	90	0	0.05	0.83	1.32	56.7	30.77	0.01	0.33	0.78	0.030303
DCMODERN	NC009567Haemophilusinfluenzae PittGG	15518	1545	1469	0.03	0.03	0.84	1.2	53.21	24.2	0.03	0.16	0.67	0.1875
DCCHIMP	NC010120Neisseriameningitidis053442	6850	839	122	0.03	0.03	0.88	1.19	27.28	5.25	0	0.28	0.78	0
DCELSIDRON1	NC010120Neisseriameningitidis053442	13536	1459	995	0.29	0.33	1.57	0.53	44.82	12.85	0	0.96	0.45	0
DCELSIDRON2	NC010120Neisseriameningitidis053442	40893	9157	1227	0.28	0.32	1.72	0.49	47.39	16.59	0	0.96	0.42	0
DCSPYNEW	NC010120Neisseriameningitidis053442	3154	1761	12	0	0	1.67	0.33	45.67	12.67	0.03	0.35	0.66	0.085714
DCSPYOLD	NC010120Neisseriameningitidis053442	3394	341	163	0.2	0.07	1.01	0.99	49.77	23.22	0	0.62	0.74	0
DCMODERN	NC010120Neisseriameningitidis053442	81524	36252	34435	0.03	0.03	0.82	1.2	58.18	23.67	0	0.16	0.91	0
DCCHIMP	NC016802Corynebacteriumdiphtheriae HC02	13435	1131	118	0	0.02	0.88	1.23	29.37	6.28	0	0.25	0.79	0
DCELSIDRON1	NC016802Corynebacteriumdiphtheriae HC02	32098	1970	1275	0.17	0.26	1.63	0.53	45.47	13.66	0	0.95	0.59	0
DCELSIDRON2	NC016802Corynebacteriumdiphtheriae HC02	34010	1106	241	0.18	0.15	1.35	0.63	36.17	12.15	0	0.74	0.67	0
DCSPYNEW	NC016802Corynebacteriumdiphtheriae HC02	2278	38	4	0	0	0.67	2	37	14.24	0.07	0.4	0.61	0.175
DCSPYOLD	NC016802Corynebacteriumdiphtheriae HC02	3882	385	169	0.03	0.06	0.93	1.13	56.44	27.67	0	0.36	0.82	0
DCMODERN	NC016802Corynebacteriumdiphtheriae HC02	35074	4729	4387	0.03	0.03	0.8	1.27	47.49	16.45	0	0.2	0.85	0
DCCHIMP	NZJLCX00000000Mycobacteriumtuberculosis T85	18229	11345	712	0.02	0.01	1.07	0.92	32.64	12.04	0	0.18	0.87	0
DCELSIDRON1	NZJLCX00000000Mycobacteriumtuberculosis T85	36572	23674	12507	0.32	0.38	1.57	0.54	49.94	13.68	0	1	0.49	0
DCELSIDRON2	NZJLCX00000000Mycobacteriumtuberculosis T85	39687	25850	2997	0.33	0.31	1.66	0.51	48.24	16.74	0	0.99	0.49	0
DCSPYNEW	NZJLCX00000000Mycobacteriumtuberculosis T85	2380	1284	21	0.29	0.38	1.36	0.6	49.1	14.98	0.02	0.72	0.53	0.027778
DCSPYOLD	NZJLCX00000000Mycobacteriumtuberculosis T85	4086	2566	1014	0.15	0.1	1.17	0.79	54.3	19.81	0	0.96	0.67	0
DCMODERN	NZJLCX00000000Mycobacteriumtuberculosis T85	21312	14822	12933	0.04	0.04	0.84	1.19	54.24	20.08	0	0.21	0.88	0
DCCHIMP	NC018828Bordetellaparapertussis Bpp5	11236	2630	482	0.02	0.01	0.91	1.11	26.5	2.94	0	0.2	0.87	0
DCELSIDRON1	NC018828Bordetellaparapertussis Bpp5	16927	2409	1677	0.2	0.19	1.47	0.61	41.18	12.64	0	0.95	0.65	0
DCELSIDRON2	NC018828Bordetellaparapertussis Bpp5	40694	5507	1073	0.21	0.21	1.47	0.63	39.6	13.43	0	0.93	0.63	0
DCSPYNEW	NC018828Bordetellaparapertussis Bpp5	2055	781	18	0	0	1	1	32.94	8.39	0.02	0.35	0.7	0.057143
DCSPYOLD	NC018828Bordetellaparapertussis Bpp5	4516	703	373	0.05	0.02	0.84	1.26	42.26	18.34	0	0.79	0.79	0
DCMODERN	NC018828Bordetellaparapertussis Bpp5	28863	6944	6614	0.03	0.02	0.78	1.27	57.77	28.46	0	0.15	0.92	0

Table S13 – MapDamage analysis was performed on the reads mapping an ancient Neandertal mitochondrial genome (1253). The number of reads mapping to each species reference genome at a quality threshold of Q30 and unduplicated is shown, along with the percentage of damage associated mutations and average read length.

Ancient Library	Reference	All Mapped Reads	Mapped Reads >Q30	Deduplicted Mapped Reads	5' C-T Proportion	3' G-A Proportion	Increase Purines Before Start	Decrease Pyrimindines Before Start	Average Length	St Dev Length	Cytosine Deamination in Dbl Strds (DeltaD)	Cytosine Deamination in Single Strds (DeltaS)	Avg Length of Single Strd Overhangs (Lambda)	Ratio of DeltaD to DeltaS	Total Reads	Percent Total Reads
DCCHIMP	ElSidron1253mtDNAgenome	12	8	1	0	0	0		36		0.14	0.46	0.54	0.304	17575167	0.00007
DCELSIDRON1	ElSidron1253mtDNAgenome	14	14	5	0	0	0.5	3	82.2	22.64	0.07	0.48	0.55	0.146	50238935	0.00003
DCELSIDRON2	ElSidron1253mtDNAgenome	0	0	0	0	0				0	0.5	0.49	0.5	1.020	48231792	0.00000
DCSPYNEW	ElSidron1253mtDNAgenome	0	0	0	0	0				0	0.5	0.5	0.51	1.000	17604340	0.00000
DCSPYOLD	ElSidron1253mtDNAgenome	185	180	56	0.18	0	0.94	1.08	82.36	25.02	0.01	0.41	0.77	0.024	13530309	0.00137
DCMODERN	ElSidron1253mtDNAgenome	2430	2397	2259	0.02	0.01	0.86	1.16	67.4	27.79	0.01	0.19	0.78	0.053	29466684	0.00825

Figure S23 – Reads rom El Sidron 2 were mapped onto a set of *Neisseria* core genes, and the resulting DNA fragments were aligned in MUGSY and compared with RAxML and bootstrapped with 100 iterations.

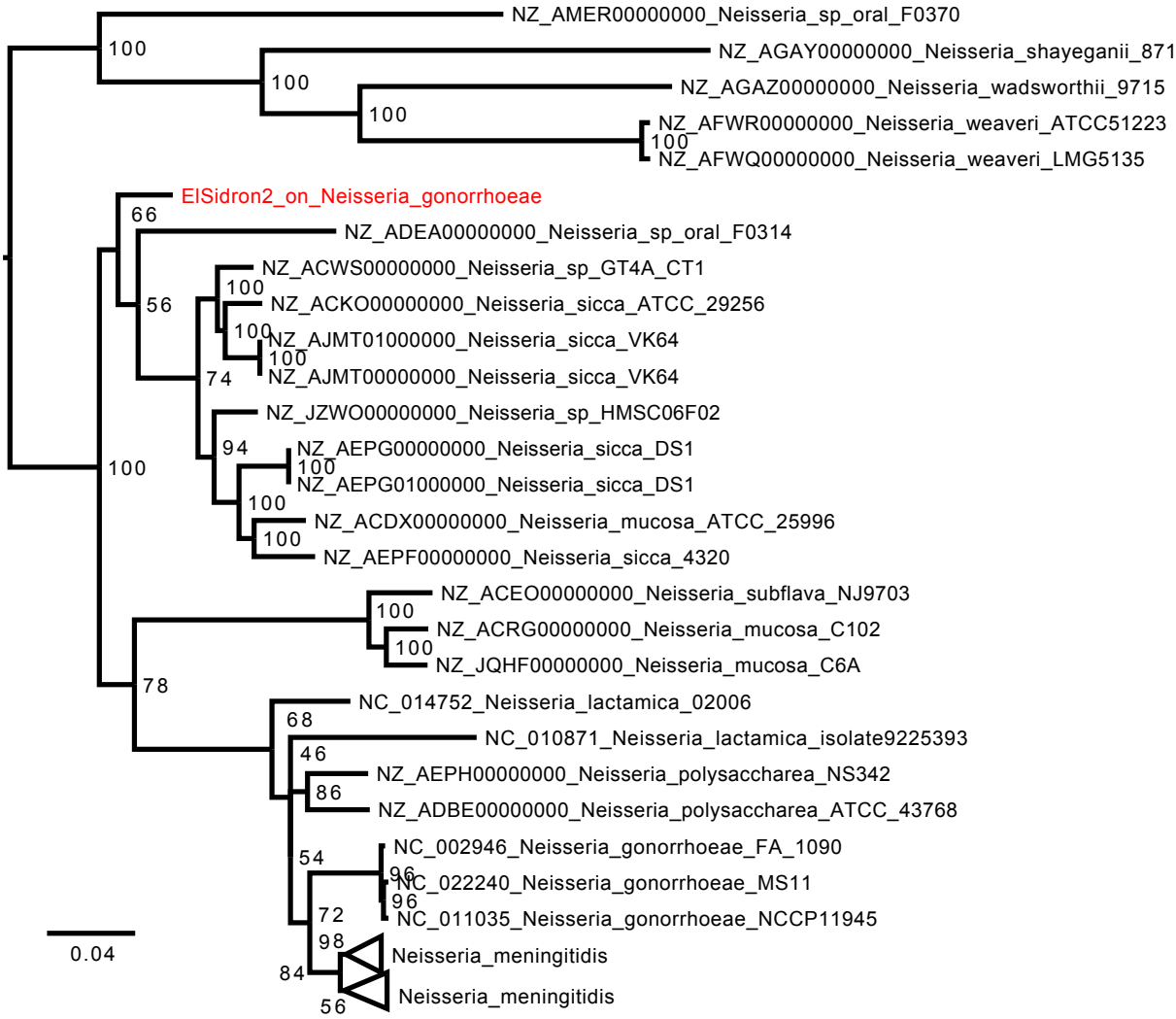


Table S14 – Filtered 16S amplicon datasets were compared to filtered shotgun datasets at the phyla level, and the proportion differences between the two datasets are shown. Values >5% are highlighted in red.

Proportion differences in filtered shotgun vs 16S rRNA data sets	Chimpanzee	El Sidron 1	El Sidron 2	Spy 1	Spy 2	Modern
Bacteria; Fusobacteria	501.41	0.16	1.52	0.01	0.00	0.31
Bacteria; Bacteroidetes	7.81	0.40	0.94	5.48	0.49	3.01
Bacteria; Actinobacteri	4.93	7.37	0.82	1.45	1.18	1.81
Bacteria; Proteobacteria	1.33	0.23	0.62	0.68	0.24	1.25
Bacteria; Synergistetes	1.15	0.99	5.47	0.01	0.00	0.32
Archaea; Euryarchaeota	0.86	10.16	17.04	0.55	53.17	1.22
Bacteria; Firmicutes	0.38	3.02	33.71	0.85	157.96	0.46
Bacteria; Chloroflexi	0.02	0.63	1.96	0.12	0.01	0.30
Bacteria; Cyanobacteria	0.01	0.00	0.00	0.00	20.14	0.00
Bacteria; Acidobacteria	0.00	0.01	0.00	0.06	0.00	0.00
Bacteria; Planctomycetes	0.00	0.00	0.00	0.06	1.14	0.00
Bacteria; Spirochaetes	0.00	1.58	9.25	0.00	0.00	3.11
Bacteria; Chlamydiae	0.00	0.00	0.15	0.00	0.00	0.00
Bacteria; Tenericutes	0.00	0.00	0.00	0.00	0.00	1.99
Bacteria; Elusimicrobia	0.00	0.00	0.13	0.00	0.00	0.00

Figure S24 – Proportions of bacterial phyla from filtered and unfiltered 16S amplicon and shotgun datasets are plotted. Samples in blue are from shotgun datasets, while red points are from 16S amplicon datasets. The different shapes of each data point correspond to the microbial phyla, which is displayed next to each phyla grouping (*e.g.* ✖ represents Proteobacteria for both 16S and shotgun data sets).

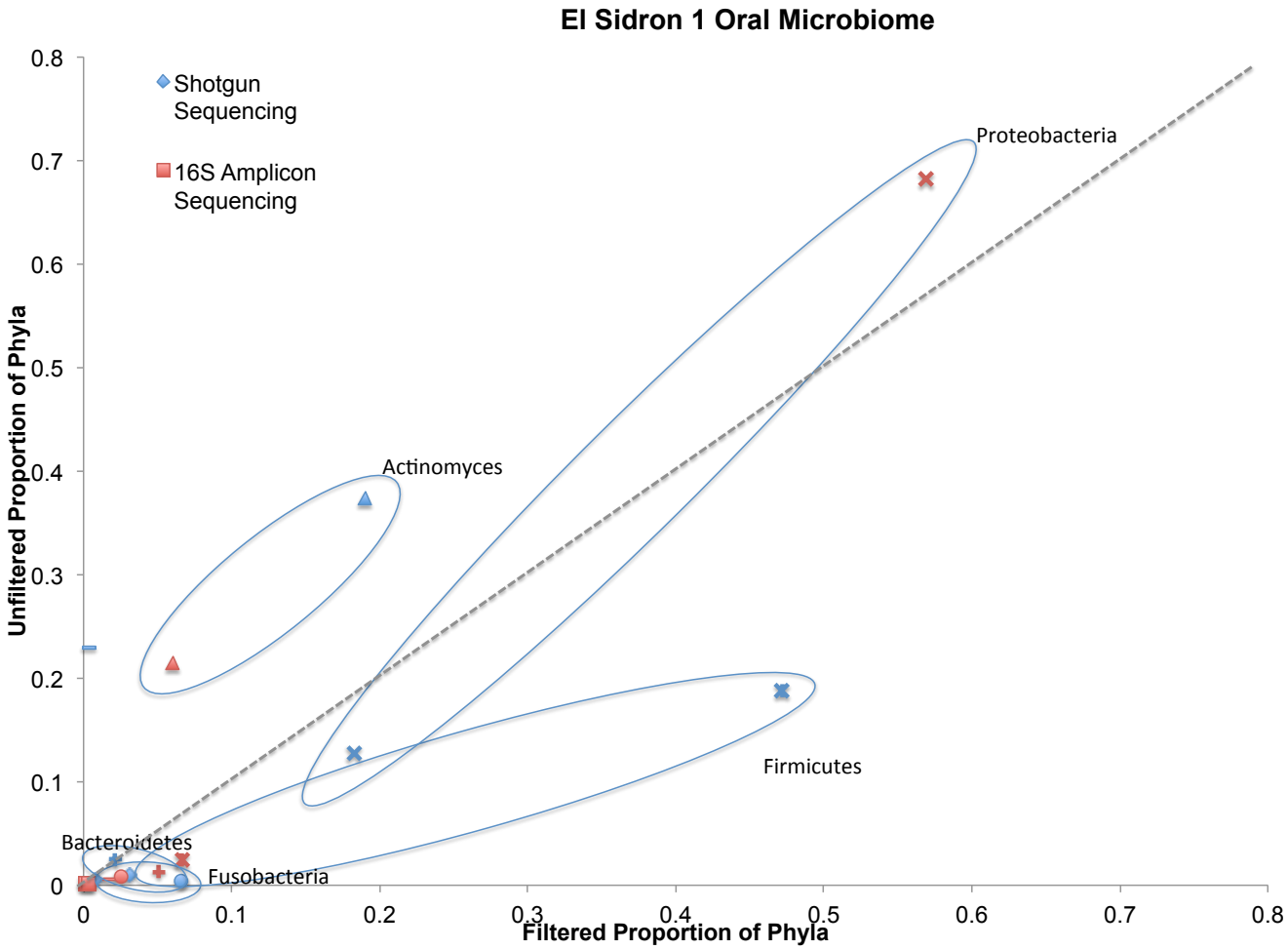


Figure S25 – Proportions of bacterial phyla from filtered and unfiltered 16S amplicon and shotgun datasets are plotted. Samples in blue are from shotgun datasets, while red points are from 16S amplicon datasets. The different shapes of each data point correspond to the microbial phyla, which is displayed next to each phyla grouping (*e.g.* ✖ represents Proteobacteria for both 16S and shotgun data sets).

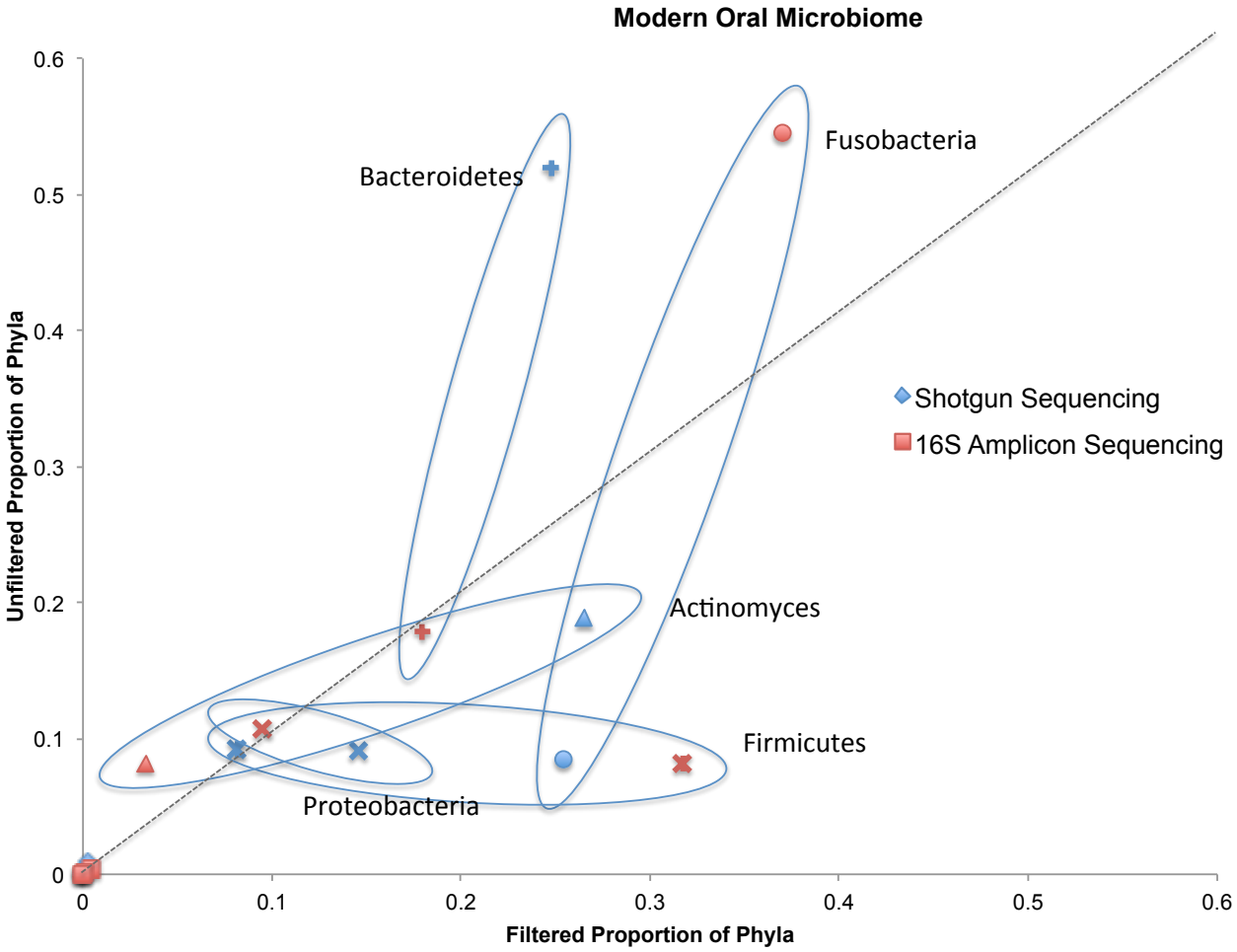


Figure S26 – Phylogenetic analysis of whooping cough in Neandertals. Shared genomic regions within publically available *Bordetella* genomes were compared to ancient *Bordetella* reads from El Sidron Neandertals using RAxML with 1,000 iterations (bootstrap values).

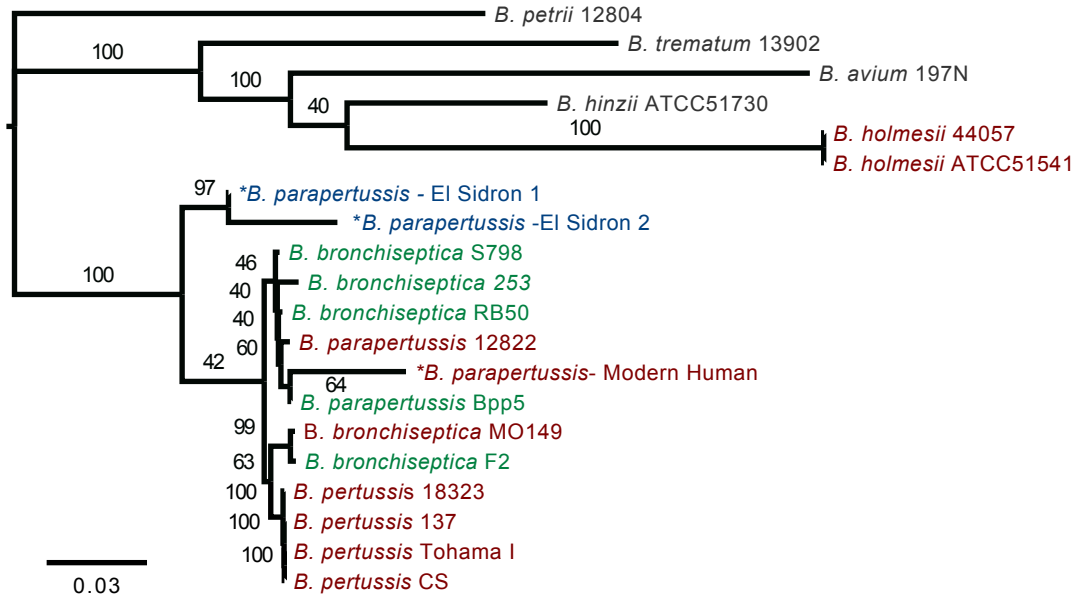


Table S15 – Mapped reads against a variety of genomes (*B. parapertussis*, *B. petrii*, *Neisseria gonorrhea*, and *Streptococcus oralis*) were compared using a script developed to compare the mapping quality scores of reads identified in two datasets to be compared directly. 'Okay' refers the number of reads that more accurate map to the first test genome (*i.e.* the genome of interest), while 'NonInf' refers to the reads that are non informative, *i.e.* do not map better to either genome. 'Contam' sequences are those that map better to the contaminant genome, *i.e.* the second genome in the test.

Sample	Read Length	Test	Okay	NonInf	Contam	Error	Proportion Bpp
ELSIDRON1	30	Bpp vs Bpet	188	0	681	0	0.216340621
ELSIDRON1	25	Bpp vs Bpet	224	0	951	0	0.190638298
ELSIDRON1	25	Bpp vs Ngon	15	0	1075	0	0.013761468
ELSIDRON1	25	Bpp vs Strep	1	0	6510	0	0.000153586

Table S16 – Coding DNA sequences (CDS) that were missing in *M. oralis neandertalensis* but present in *M. oralis* in modern humans are displayed, only if the function was known. 81 genes missing in *M. oralis neandertalensis* were annotated as hypothetical proteins and are not displayed.

Name	Minimum	Maximum	Length	Direction
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase CDS	144007	144702	696	reverse
2,5-diketo-D-gluconic acid reductase CDS	24887	25720	834	forward
2,5-diketo-D-gluconic acid reductase CDS	26333	27230	898	forward
3-phosphoshikimate 1-carboxyvinyltransferase CDS	348407	349723	1317	reverse
7-cyano-7-deazaguanine synthase CDS	146022	146696	675	forward
ABC transporter CDS	2247	3892	1646	reverse
ABC transporter CDS	213147	213917	771	reverse
ABC transporter CDS	22560	24290	1731	reverse
acetyltransferase CDS	217914	218720	807	forward
alpha/beta hydrolase CDS	15018	15992	975	reverse
ammonia channel protein CDS	<324463	325557	>1095	reverse
appr-1-p processing enzyme family domain-containing protei	421943	422725	783	reverse
asparagine synthetase CDS	274918	276366	1449	reverse
ATPase CDS	95405	96640	1236	reverse
ATPase CDS	99167	100366	1200	reverse
ATPase CDS	27818	29050	1233	reverse
cation transporter CDS	182373	184227	1855	forward
cation transporter CDS	16062	18081	2020	reverse
CRISPR-associated endonuclease Cas2 CDS	27409	27675	267	reverse
CRISPR-associated endoribonuclease Cas6 CDS	124963	125703	741	reverse
CRISPR-associated protein CDS	18877	20124	1248	reverse
cupin CDS	25795	26226	432	forward
cysteine desulfurase CDS	341274	342458	1185	reverse
cysteine synthase CDS	346747	347693	947	reverse
DNA methyltransferase CDS	117589	119136	1548	forward
DNA mismatch repair protein MutT CDS	265100	265504	405	forward
DUF2634 domain-containing protein CDS	71266	71694	429	reverse
endonuclease III CDS	347784	348407	624	reverse
formate dehydrogenase family accessory protein FdhD CDS	393696	394460	765	forward
glycerol-3-phosphate dehydrogenase CDS	144712	145680	969	reverse
GshA CDS	167178	168571	1394	forward
HNH endonuclease CDS	97246	97863	618	reverse
integrase CDS	110585	111695	1111	reverse
iron-sulfur cluster assembly scaffold protein NifU CDS	340890	341261	372	reverse
KlaA protein CDS	52163	53302	1140	reverse
KTSC domain-containing protein CDS	101752	101964	213	forward
MATE family efflux transporter CDS	263908	265272	1365	reverse
membrane protein CDS	90942	91751	810	forward
methyltransferase CDS	350273	351055	783	reverse
nitrogen regulatory protein P-II 1 CDS	323223	323561	339	reverse
NrdH-redoxin CDS	27815	28096	282	forward
NUDIX domain-containing protein CDS	146181	146597	417	forward
phosphate ABC transporter ATPase CDS	211638	212692	1055	reverse
phosphatidate cytidylyltransferase CDS	410238	410912	675	reverse
PIN domain-containing protein CDS	126213	126602	390	forward
pyridoxamine 5'-phosphate oxidase CDS	18405	18806	402	reverse
QacE family quaternary ammonium compound efflux SMR tr	145813	146133	321	forward
restriction endonuclease CDS	190474	190989	516	reverse
restriction endonuclease CDS	105010	106892	1883	forward
SAM-dependent methyltransferase CDS	217129	217911	783	forward
SAM-dependent methyltransferase CDS	220837	221415	579	forward
Sir2 silent information regulator family NAD-dependent deac	421093	421941	849	reverse
sugar fermentation stimulation protein CDS	447396	448158	763	forward
type I restriction endonuclease CDS	114455	117580	3126	forward